

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 19:39:52 ; Search time 3754 Seconds  
(without alignments)  
11341.877 Million cell updates/sec

Title: US-09-805-311-1  
Perfect score: 1463  
Sequence: 1 caccagaatactgcggcc.....aaaaaaaaaaaaaaaaaaaa 1463

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2034640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pi:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rnd:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htg\_hum:\*
- 40: em\_htg\_mus:\*
- 41: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1463	100.0	1463	6	ARI52403	Sequence ARI52403
2	1421.8	97.2	1478	6	ARI52406	Sequence ARI52406
3	1409.8	96.4	1341	6	ARI52404	Sequence ARI52404
4	1350.2	92.3	1381	6	ARI52405	Sequence ARI52405
5	927.4	63.4	1354	8	AB021666	AB021666
6	326.8	22.3	1469	5	XL068141	U68141 xenopus lae
7	308.6	21.1	1461	5	XL064563	U64563 xenopus lae
8	307	21.0	1357	5	AF065397	AF065397 xenopus l
9	305.4	20.9	1149	5	AF036327	AF036327 xenopus l
10	305.4	20.9	2163	10	BC010203	BC010203 Mus muscu
11	305.4	20.9	163475	10	AC026761	AC026761 Mus muscu
12	303.8	20.8	2102	10	BC027295	BC027295 Mus muscu
13	302.8	20.7	1757	9	HSRAD2	X76771 Homo sapien
14	302.8	20.7	2031	9	BC000323	BC000323 Homo sapi
15	302.8	20.7	6536	9	AF523117	AF523117 Homo sapi
16	302.8	20.7	13827	2	AP000591	AP000591 Homo sapi
17	302.8	20.7	169053	2	AP002380	AP002380 Homo sapi
18	302.8	20.7	185035	9	AC004770	AC004770 Homo sapi
19	302.8	20.7	196080	9	AC004228	AC004228 Homo sapi
20	302.6	20.7	1144	9	HUMFENIA	L37374 Homo sapien
21	299	20.4	1146	5	AB058602	AB058602 Gallus ga
22	292.4	20.0	39563	9	HSU73629	U73629 Human chrom
23	288.2	19.7	2192	10	AF281018	AF281018 Rattus no
24	288.2	19.7	182315	2	AC127862	AC127862 Rattus no
25	282.6	19.3	114000	2	AC125412	AC125412 Leishmani
26	279.8	19.1	2033	10	MUSFENIX	L26320 Mouse flap
27	278.2	19.0	6361	10	AY014962	AY014962 Mus muscu
28	274.6	18.8	2114	8	SCYKL113C	Z28113 S.cerevisia
29	274.6	18.8	10660	8	S93804	S93804 BAF1/ABF1/Y
30	259.4	17.7	2299	8	SPRAD2	X77041 Saccharomyc
31	259.4	17.7	28491	8	SPAC3G6	Z99167 S.pombe chr
32	259	17.7	151865	2	AL356236	AL356236 Homo sapi
33	259	17.7	179461	2	AC019177	AC019177 Homo sapi
34	257.4	17.6	173256	2	AC073301	AC073301 Homo sapi
35	257.4	17.6	178806	2	AC013557	AC013557 Homo sapi
36	252.6	17.3	37688	2	AL513222	AL513222 Drosophil
37	252.6	17.3	37688	3	DMC5G0003	AL031863 Drosophil
38	252.6	17.3	95850	2	AC020063	AC020063 Drosophil
39	252.6	17.3	154985	3	AC004641	AC004641 Drosophil
40	252.6	17.3	160338	3	AC099021	AC099021 Drosophil
41	252.6	17.3	179283	3	AC099030	AC099030 Drosophil
42	252.6	17.3	272948	3	AE003805	AE003805 Drosophil
43	241.8	16.5	194322	3	AC04791	AC04791 Caenorhab
44	241.8	16.5	319857	2	AC006782	AC006782 Caenorhab
45	237	16.2	2783	3	AF093702	AF093702 Plasmodiu

ALIGNMENTS

RESULT 1	ARI52403	Sequence 1 from patent US 6232527.	1463 bp	DNA	Linear	PAT 08-AUG-2001
LOCUS	ARI52403	Sequence 1 from patent US 6232527.				
DEFINITION	ARI52403	Sequence 1 from patent US 6232527.				
ACCESSION	ARI52403	Sequence 1 from patent US 6232527.				
VERSION	ARI52403.1	GI:15118453				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1463)					
AUTHORS	Mahajan, P.B.					
TITLE	Maize Rad2/FEN-1 orthologues and uses thereof					
JOURNAL	Patent: US 6232527-A 1 15-MAY-2001;					
FEATURES	Location/Qualifiers					

```

source      1. 1463
/organism="unknown"
BASE COUNT  466 a 292 c 361 g 344 t
ORIGIN

Query Match      100.0%; Score 1463; DB 6; Length 1463;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACGAGAAATAGCTCGCGCGCGGTTCTTGGCCACTCCGGCTCAGCGCGCGCGCGCA 60
   |||||||
Db 1 CACGAGAAATAGCTCGCGCGCGGTTCTTGGCCACTCCGGCTCAGCGCGCGCGCGCA 60

QY 61 CCCGCCACAGCGCGCGAGAGATGGGATCAAGGGTTTCAGAAACTGCTGGCGGAC 120
   |||||||
Db 61 CCCGCCACAGCGCGCGAGAGATGGGATCAAGGGTTTCAGAAACTGCTGGCGGAC 120

QY 121 AATGCGCCAAAGGGGATGAAGGAGCAGAAATTCGAGAGCTACTTCGGCGCGCAAAATCGCC 180
   |||||||
Db 121 AATGCGCCAAAGGGGATGAAGGAGCAGAAATTCGAGAGCTACTTCGGCGCGCAAAATCGCC 180

QY 181 GTCGAGCGCCAGCATGAGCATATACAGTTCCTGATTGTGTTGGAAGGACAGGCGATGAA 240
   |||||||
Db 181 GTCGAGCGCCAGCATGAGCATATACAGTTCCTGATTGTGTTGGAAGGACAGGCGATGAA 240

QY 241 ACTCTCAAAATGAAGCTGGTGAAGTCACTAGTCACTTTCGAGGAATGTTCAACCGGACA 300
   |||||||
Db 241 ACTCTCAAAATGAAGCTGGTGAAGTCACTAGTCACTTTCGAGGAATGTTCAACCGGACA 300

QY 301 ATAAGATTACTGGAAGGGGAATCAAGCCAGTTTATGTTTGTATGGCAAGCTTCCTGAT 360
   |||||||
Db 301 ATAAGATTACTGGAAGGGGAATCAAGCCAGTTTATGTTTGTATGGCAAGCTTCCTGAT 360

QY 361 ATGAAGAAACAGAGCTTGCTAAAGATACCTCAAAAGAGATGATCAACCAAGATCTG 420
   |||||||
Db 361 ATGAAGAAACAGAGCTTGCTAAAGATACCTCAAAAGAGATGATCAACCAAGATCTG 420

QY 421 ACTGAGCAGTAGAGTAGAGATGAAGATCGGATTCGAAATGAGCAGAGAGACTGTA 480
   |||||||
Db 421 ACTGAGCAGTAGAGTAGAGATGAAGATCGGATTCGAAATGAGCAGAGAGACTGTA 480

QY 481 AAGTCAACAGGCAACACAGAGATTGTAACGGCTATTAGACTTATGGGGTTCCCT 540
   |||||||
Db 481 AAGTCAACAGGCAACACAGAGATTGTAACGGCTATTAGACTTATGGGGTTCCCT 540

QY 541 GTTGTAGAGCACCCTCTGAGCAGAGAGAGATGTCGACCCCTTTCATTAACGATAAG 600
   |||||||
Db 541 GTTGTAGAGCACCCTCTGAGCAGAGAGAGATGTCGACCCCTTTCATTAACGATAAG 600

QY 601 GTGTTCCGCTGTGCTTCAGAGATATGGACTCCCTTACTTTTGGGGCTCCACGGTTCCCT 660
   |||||||
Db 601 GTGTTCCGCTGTGCTTCAGAGATATGGACTCCCTTACTTTTGGGGCTCCACGGTTCCCT 660

QY 661 CGTCATTTAATGATCCAAAGTTCAGAGAAATACCTGTGATGGAATTTGATGTTGCAAG 720
   |||||||
Db 661 CGTCATTTAATGATCCAAAGTTCAGAGAAATACCTGTGATGGAATTTGATGTTGCAAG 720

QY 721 GTTTTGGAGGAGCTTGAACCTCAGAGAGAGATGTCGACCCCTTTCATTTGTCATGTTGGA 780
   |||||||
Db 721 GTTTTGGAGGAGCTTGAACCTCAGAGAGAGATGTCGACCCCTTTCATTTGTCATGTTGGA 780

QY 781 TGTGACTATTGTGATAGCATCAAGGATATCGGGGGCAACACACTCTGAAACTTATTTCGT 840
   |||||||
Db 781 TGTGACTATTGTGATAGCATCAAGGATATCGGGGGCAACACACTCTGAAACTTATTTCGT 840

QY 841 CAACATGGGTCCATAGAAAGCATCTTGGAGATCTTTAATAAGACAGATATCAAAATTCCT 900
   |||||||
Db 841 CAACATGGGTCCATAGAAAGCATCTTGGAGATCTTTAATAAGACAGATATCAAAATTCCT 900

QY 901 GAGGACTGGCCTTTACCAAGAGCTGCACGCTTGTTCAGAGAGCCTAATGTCACATTGGAT 960
   |||||||
Db 901 GAGGACTGGCCTTTACCAAGAGCTGCACGCTTGTTCAGAGAGCCTAATGTCACATTGGAT 960

```

```

QY 961 ATTCTGAGCTAAATGAGTGCACCTGATGAGGAGGCTCTCATAGTTTCTCTGTAATA 1020
   |||||||
Db 961 ATTCTGAGCTAAATGAGTGCACCTGATGAGGAGGCTCTCATAGTTTCTCTGTAATA 1020

QY 1021 GATAATGTTTCAACGAGAGATCGGTTGACAAAGGCCATAGAGAGATCAAAATCTGCCAAG 1080
   |||||||
Db 1021 GATAATGTTTCAACGAGAGATCGGTTGACAAAGGCCATAGAGAGATCAAAATCTGCCAAG 1080

QY 1081 AATAAATCTCGCAAGGAAGACTCGAGTCCCTTTTCAAGCCCAACTGCCACACATCAGCA 1140
   |||||||
Db 1081 AATAAATCTCGCAAGGAAGACTCGAGTCCCTTTTCAAGCCCAACTGCCACACATCAGCA 1140

QY 1141 CCGCTAAACCGAAGGAGACTTCGGATATAAACAAGCAGGAGCTGCGAACAAGAAAAACA 1200
   |||||||
Db 1141 CCGCTAAACCGAAGGAGACTTCGGATATAAACAAGCAGGAGCTGCGAACAAGAAAAACA 1200

QY 1201 AAGGCTGGTGAAGAGAAATAATCTTGGATGCTTGGATGCTTGGATGCTTGGATGCTTGGATGCT 1260
   |||||||
Db 1201 AAGGCTGGTGAAGAGAAATAATCTTGGATGCTTGGATGCTTGGATGCTTGGATGCTTGGATGCT 1260

QY 1261 CAGCGGTGGCTGATCACTTCGCTTAGATTTAATACTCCCTGTTTAACTCAGAGCTTT 1320
   |||||||
Db 1261 CAGCGGTGGCTGATCACTTCGCTTAGATTTAATACTCCCTGTTTAACTCAGAGCTTT 1320

QY 1321 GGTAAAAATTTGCTCATCTGTTTCAAGCTGGGTAAGTTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1380
   |||||||
Db 1321 GGTAAAAATTTGCTCATCTGTTTCAAGCTGGGTAAGTTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1380

QY 1381 TACCAAGTAAACAACATATCGCTGTTTCTTACTTCTTGTGCTTGTGAAGTAAACAAAAA 1440
   |||||||
Db 1381 TACCAAGTAAACAACATATCGCTGTTTCTTACTTCTTGTGCTTGTGAAGTAAACAAAAA 1440

QY 1441 AAAAAAAGGAAAAAAGAAAAA 1463
   |||||||
Db 1441 AAAAAAAGGAAAAAAGAAAAA 1463

```

```

RESULT 2
AR152406 LOCUS 1478 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 7 from patent US 6232527.
ACCESSION AR152406
VERSION AR152406.1 GI:15118456
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1478)
AUTHORS Mahajan, P.B.
TITLE Maize Rad2/FEN-1 orthologues and uses thereof
JOURNAL Patent: US 6232527-A 7 15-MAY-2001;
FEATURES
    Location/Qualifiers
        source
            1..1478
                /organism="unknown"
BASE COUNT 463 a 302 c 365 g 348 t
ORIGIN

```

```

Query Match      97.2%; Score 1421.8; DB 6; Length 1478;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 1435; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 7 AATAGCTCGCGCGCGGTTCTTTCGCGCACCTCCGGCTCAGCGCGCGCGCGCCGCC 66
   |||||||
Db 19 AATAGCTCGCGCGCGGTTCTTTCGCGCACCTCCGGCTCAGCGCGCGCGCGCCGCC 78

QY 67 ACAGCGCGCGCAGACGAGATGGGCATCAAGGGTTTCACGAAACTGCTGCGGACATGCG 126
   |||||||
Db 79 ACAGCGCGCGCAGACGAGATGGGCATCAAGGGTTTCACGAAACTGCTGCGGACATGCG 138

QY 127 CCCAAGCGGATGAAGAGCAGAGTTCGAGAGTACTTCGCGCGCAAAATCGCGCTCGAC 186
   |||||||
Db 139 CCCAAGCGGATGAAGAGCAGAGTTCGAGAGTACTTCGCGCGCAAAATCGCGCTCGAC 198

```

QY	187	GCAGCATGAGCATATACCAAGTTCCTGATTTAGTTGGAAAGACAGGCATGGAAACTCTC	246
Db	199	GCCAGCATGAGCATCTACCAAGTTCTCTGATAGTAGTTTGGAAAGACAGGCATGGAACTCTC	258
QY	247	ACAAATGAAGCTGGTGAAGTCACTAGTCATTGCAAGGAATGTTCAACGGCAATAAGA	306
Db	259	ACAAATGAAGCTGGTGAAGTCACTAGTCATTGCAAGGAATGTTCAACGGCAATAAGA	318
QY	307	TTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTCATGGCAAGCCTCCTGATATGAAG	366
Db	319	TTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTCATGGCAAGCCTCCTGATATGAAG	378
QY	367	AAACAAGAGCTTGTCTAAAGATACCTCAAAAAGAGATGATGCAACCAAGATCTGACTGAG	426
Db	379	AAACAAGAGCTTGTCTAAAGATACCTCAAAAAGAGATGATGCAACCAAGATCTGACTGAG	438
QY	427	GCAGTAGAGGTAGGAGATAAGATCCGATTGAAAATTTGACGAAGAGGACTGTAAGGTC	486
Db	439	GCAGTAGAGGTAGGAGATAAGATCCGATTGAAAATTTGACGAAGAGGACTGTAAGGTC	498
QY	487	ACAAGGCAACAACAAGAGATTGTAAGCGCTATTAAAGACTTATGGGGGTTCCTGTTGTA	546
Db	499	ACAAGGCAACAACAAGAGATTGTAAGCGCTATTAAAGACTTATGGGGGTTCCTGTTGTA	558
QY	547	GAGGCACCTTCTGAAGCAGAGACAGAAATGTGCGACCTTTTGATATAACGATAGGTGTTTC	606
Db	559	GAGGCACCTTCTGAAGCAGAGACAGAAATGTGCGACCTTTTGATATAACGATAGGTGTTTC	618
QY	607	GCYTTGCTTCAGAGATATGACATCCCTTACTTTTGGGGCTCCACGGTTCCTTCGTCAT	666
Db	619	GCYTTGCTTCAGAGATATGACATCCCTTACTTTTGGGGCTCCACGGTTCCTTCGTCAT	678
QY	667	TTAATGGATCCAAGTTCCAAAGAAATACCTGTGATGGAATTTGATTTGCCAAGGTTTG	726
Db	679	TTAATGGATCCAAGTTCCAAAGAAATACCTGTGATGGAATTTGATTTGCCAAGGTTTG	738
QY	727	GAGGAGCTTGAATCTACCATGAGACAGTTCATTTGATTTGTGTCATCCTGTGTGATGTGAC	786
Db	739	GAGGAGCTTGAATCTACCATGAGACAGTTCATTTGATTTGTGTCATCCTGTGTGATGTGAC	798
QY	787	TATTGTATAGCATCAAAGGTATCGGGGGCAACAGCTCTGAACTTATTCCTCAACAT	846
Db	799	TATTGTATAGCATCAAAGGTATCGGGGGCAACAGCTCTGAACTTATTCCTCAACAT	858
QY	847	GGGTCCATAGAAAGCATCTTGGAGATCTTATAAAGACAGATATCAAAATCCTGAGGAC	906
Db	859	GGGTCCATAGAAAGCATCTTGGAGATCTTATAAAGACAGATATCAAAATCCTGAGGAC	918
QY	907	TGGCCTTTACCAAGAAGCTCGAGCTTGTTCAGGAGCCATAATGTCATATGGATATTCCT	966
Db	919	TGGCCTTTACCAAGAAGCTCGAGCTTGTTCAGGAGCCATAATGTCATATGGATATTCCT	978
QY	967	GAGCTAAATTTGGACTGCACCTGATGAGGAGGCTCTATAAGTTTCCCTGGTAAAGATAAT	1026
Db	979	GAGCTAAATTTGGACTGCACCTGATGAGGAGGCTCTATAAGTTTCCCTGGTAAAGATAAT	1038
QY	1027	GTTTTCACGAAGATCGGTTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAATAAA	1086
Db	1039	GTTTTCACGAAGATCGGTTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAATAAA	1098
QY	1087	TCGTGCAAGGAAGACTTCGAGTCTCTTTTTCAGGCCAATGCCACACATCAGCACCGCTA	1146
Db	1099	TCGTGCAAGGAAGACTTCGAGTCTCTCTTTTTCAGGCCAATGCCACACATCAGCACCGCTA	1158
QY	1147	AAACGGAAGGAGACTTCGGATATAACACAGCAGGCGCTGCGNACAGAAACAAAGGCT	1206
Db	1159	AAACGGAAGGAGACTTCGGATATAACACAGCAGGCGCTGCGNACAGAAACAAAGGCT	1218
QY	1207	GGTGGAAAGAGAAATAATCTTCGATGCTTGATGTACAACTACGACTACGAAAGACGCGG	1266
Db	1219	GGTGGAAAGAGAAATAATCTTCGATGCTTGATGTACAACTACGACTACGAAAGACGCGG	1278
QY	1267	TGGCGTGATCACCTTCGCTTAGATTTATTTAACTCCCTGTTTTTAACCTACGAGGTTTGGTAAA	1326

Db	1279	TGGCATGATCACTTCGCCCTAGATTATTAACTCCCTGTTTAACTCAGACCTTTGGTGCA	1338
Qy	1327	AGTTTGGTTCATGTTTCAAGCTGGGTAAGTTAGTTGTTTGAAGAGATTGGTGTACCAA	1386
Db	1339	AGTTTGGCCATGTTTCAAGCTGGGTAAGTTAGTTGTTTGAAGAGATTGGTGTACCAA	1398
Qy	1387	GTAACAAACTATCGCTGTTTGTACCTCTGTCCTTTGAAGTAAAAAATAAAAAA	1446
Db	1399	GTAACAAACTATCGCTGTTTGTACCTCTGTCCTTTGAAGTAAAAAATAAAAAA	1458
Qy	1447	AAAAAATAAAAAAATAAAAAA	1463
Db	1459	AAAAAATAAAAAAATAAAAAA	1475
<p>RESULT 3</p> <p>LOCUS ARI52404 1541 bp DNA linear PAT 08-AUG-2001</p> <p>DEFINITION Sequence 3 from patent US 6232527.</p> <p>ACCESSION ARI52404</p> <p>VERSION ARI52404.1 GI:15118454</p> <p>KEYWORDS</p> <p>SOURCE Unknown.</p> <p>ORGANISM Unclassified.</p> <p>REFERENCE 1 (bases 1 to 1541)</p> <p>AUTHORS Mahajan, P. B.</p> <p>TITLE Maize Rad2/FEN-1 orthologues and uses thereof</p> <p>JOURNAL Patent: US 6232527-A 3 15-MAY-2001;</p> <p>FEATURES</p> <p>source Location/Qualifiers</p> <p>1..1541</p> <p>/organism="unknown"</p>			
BASE COUNT	473 a	308 c	377 g
ORIGIN			
<p>Query Match 96.4%; Score 1409.8; DB 6; Length 1541;</p> <p>Best Local Similarity 99.9%; Pred. No. 0;</p> <p>Matches 1411; Conservative 0; Mismatches 2; Indels 0; Gaps 0</p>			
Qy	19	CGCGGTTCTTTCGGCCACTTCGGCTCAGCGCGCGCCGCCACCGCCACAGCC	78
Db	13	CGCGGTTCTTTCGGCCACTTCGGCTCAGCGCGCGCCGCCACCGCCACAGCC	72
Qy	79	GACGAGATGGGCATCAAGGGTTTGACAAACTGCTGGCGGCAATGCGCCCAAGGCGATG	138
Db	73	GACGAGATGGGCATCAAGGGTTTGACAAACTGCTGGCGGCAATGCGCCCAAGGCGATG	132
Qy	139	AGGAGCACAAGTTCGAGAGCTTCTCGCGCGCAAAATCGCGTCGACCGCCAGCATGAGC	198
Db	133	AGGAGCACAAGTTCGAGAGCTTCTCGCGCGCAAAATCGCGTCGACCGCCAGCATGAGC	192
Qy	199	ATATACCACTTCCTGATTCTAGTTTGAAGGACAGGCATGGAACCTCTCAAAATGAAGCT	258
Db	193	ATATACCACTTCCTGATTCTAGTTTGAAGGACAGGCATGGAACCTCTCAAAATGAAGCT	252
Qy	259	GGTGAAGTCACATAGTCATTTGCAAGGAATGTTCAACCGGCAATATAGATTACTGGAAGCG	318
Db	253	GGTGAAGTCACATAGTCATTTGCAAGGAATGTTCAACCGGCAATATAGATTACTGGAAGCG	312
Qy	319	GGAATCAAGCCAGTTTATGTTTTTGTATGTCAGGCCCTCTGATATGAAGAAACAAGAGCTT	378
Db	313	GGAATCAAGCCAGTTTATGTTTTTGTATGTCAGGCCCTCTGATATGAAGAAACAAGAGCTT	372
Qy	379	GCTAAAGATACTCAAAAAGAGATGATGCAACAAAGATCTGACTGAGGCGATGAGGTA	438
Db	373	GCTAAAGATACTCAAAAAGAGATGATGCAACAAAGATCTGACTGAGGCGATGAGGTA	432
Qy	439	GGAGATAAGATCGGATTCGAAAATTTGACAAGAGGACTGTTAAAGGTCTCAAGGCAACAC	498
Db	433	GGAGATAAGATCGGATTCGAAAATTTGACAAGAGGACTGTTAAAGGTCTCAAGGCAACAC	492
Qy	499	AACGAAGATTGTAAGCGGCTATTAAAGACTTATGGGGTTCCTGTTGTAGAGGCACCTTCT	558

Db 493 AACGAGATGTAACGGCTATTAGACTATGGGGTTCCTGTTGATAGAGCACCTTCT 552  
QY 559 GAAGCAGAGCAGATGTCAGCCCTTGCATAAACGATAGGTGTCGCTGCTTCA 618  
Db 553 GAAGCAGAGCAGATGTCAGCCCTTGCATAAACGATAGGTGTCGCTGCTTCA 612  
QY 619 GAAGATATGGACTCCCTTACTTCTTGGGGCTCCACGGTTCCTTCGTCATTTATGATCCA 678  
Db 613 GAAGATATGGACTCCCTTACTTCTTGGGGCTCCACGGTTCCTTCGTCATTTATGATCCA 672  
QY 679 AGTTCCTCAAGAAATACCTGTCATGGAATTTGATGTTGCCAAGGTTTGGAGAGCTTGA 738  
Db 673 AGTTCCTCAAGAAATACCTGTCATGGAATTTGATGTTGCCAAGGTTTGGAGAGCTTGA 732  
QY 739 CTCACCATGGACAGTTCATGATTTGTCATCTCTGTCGTCATTTATGATAGC 798  
Db 733 CTCACCATGGACAGTTCATGATTTGTCATCTCTGTCGTCATTTATGATAGC 792  
QY 799 ATCAAGATATCGGGGGGCAACAGCTCTGAACCTTATTCGTCACATGGGTCCATAGAA 858  
Db 793 ATCAAGATATCGGGGGGCAACAGCTCTGAACCTTATTCGTCACATGGGTCCATAGAA 852  
QY 859 AGCATCTTGGAGATCTTAATAAAGACAGATCAAAATTCCTGAGCTGGCTTACCAA 918  
Db 853 AGCATCTTGGAGATCTTAATAAAGACAGATCAAAATTCCTGAGCTGGCTTACCAA 912  
QY 919 GAAGCTGCACCTGTTTCAAGAGCCTAATGTCACATTTGATGATTTCTGAGCTAAAATGG 978  
Db 913 GAAGCTGCACCTGTTTCAAGAGCCTAATGTCACATTTGATGATTTCTGAGCTAAAATGG 972  
QY 979 ACTGCACCTGATGAGGAGGTCTCATAGTTTCTGTTAAAGATATGTTTCAACGAA 1038  
Db 973 ACTGCACCTGATGAGGAGGTCTCATAGTTTCTGTTAAAGATATGTTTCAACGAA 1032  
QY 1039 GATCGGTGCAAAAGGGCCATAGAGAAGATCAAAATCTGCAAGAAATAATCTGCCAAGGA 1098  
Db 1033 GATCGGTGCAAAAGGGCCATAGAGAAGATCAAAATCTGCAAGAAATAATCTGCCAAGGA 1092  
QY 1099 AGACTCGAGTCTTTTCAAGCCAACTGCCACACATCAGCACCGCTAAAACGGAAGAG 1158  
Db 1093 AGACTCGAGTCTTTTCAAGCCAACTGCCACACATCAGCACCGCTAAAACGGAAGAG 1152  
QY 1159 ACTTCGGATAAAACAGCAGGAGCTGCGAACAAGAAACAAAGGCTGTTGGAAGAG 1218  
Db 1153 ACTTCGGATAAAACAGCAGGAGCTGCGAACAAGAAACAAAGGCTGTTGGAAGAG 1212  
QY 1219 AAATAATCTTGGATGCTTGATGTACAACTACGACTACGAAAGCAGCGTGGCGTGATCAC 1278  
Db 1213 AAATAATCTTGGATGCTTGATGTACAACTACGACTACGAAAGCAGCGTGGCGTGATCAC 1272  
QY 1279 TTCGCTTAGATTAATTAACCTCCCTGTTTAACTCAGAGCTTTGGTAAAGTTTGGCTCATG 1338  
Db 1273 TTCGCTTAGATTAATTAACCTCCCTGTTTAACTCAGAGCTTTGGTAAAGTTTGGCTCATG 1332  
QY 1339 TTTCAAGCTGGGTAACTTAGTGTGTTTGAAGAGATTTGTTACCAAGTAAACAAACTT 1398  
Db 1333 TTTCAAGCTGGGTAACTTAGTGTGTTTGAAGAGATTTGTTACCAAGTAAACAAACTT 1392  
QY 1399 ATCGCTGTTTTTACTTCTTGTGCTTTGAAGTA 1431  
Db 1393 ATCGCTGTTTTTACTTCTTGTGCTTTGAAGTA 1425

RESULT 4  
AR152405  
LOCUS AR152405 1381 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 5 from patent US 6232527.  
ACCESSION AR152405  
VERSION AR152405.1 GI:15118455  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

Unclassified.  
REFERENCE 1 (bases 1 to 1381)  
AUTHORS Mahajan, P. B.  
TITLE Maize Rad2/FEN-1 orthologues and uses thereof  
JOURNAL Patent: US 6232527-A 5 15-MAY-2001;  
FEATURES Location/Qualifiers  
source 1. 1381  
BASE COUNT 441 a 269 c 346 g 325 t  
ORIGIN  
Query Match 92.3%; Score 1350.2; DB 6; Length 1381;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1355; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 49 CGCCGCCGCCACCGCCACAGCCGCCACAGAGATGGGATCAAGGGTTTGACGAAA 108  
Db 1 CGACCCACGCGCTCCGCGCCACAGCCGCCACAGAGATGGGATCAAGGGTTTGACGAAA 60  
QY 109 CTGCTGGCGGACAATGCGCCCAAGCGGATGAAGAGCAGAAAGTTTCGAGAGCTACTTCGCG 168  
Db 61 CTGCTGGCGGACAATGCGCCCAAGCGGATGAAGAGCAGAAAGTTTCGAGAGCTACTTCGCG 120  
QY 169 CGCAAAATCGCGCTCGAGCCGAGCATATACCAAGTTTCTGATGTAGTTGGAAG 228  
Db 121 CGCAAAATCGCGCTCGAGCCGAGCATATACCAAGTTTCTGATGTAGTTGGAAG 180  
QY 229 ACAGGCTGGAACCTCTCAAAATCAAGCTGCTGAAGTCACTAGTCAATTTGCAAGGAATG 288  
Db 181 ACAGGCTGGAACCTCTCAAAATCAAGCTGCTGAAGTCACTAGTCAATTTGCAAGGAATG 240  
QY 289 TTCAACCGGACAATAAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTGATGGC 348  
Db 241 TTCAACCGGACAATAAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTGATGGC 300  
QY 349 AAGCCTCTGATATGAAGAAACAGAGCTTCTTAAAGATACTCAAAAGAGATGATGCA 408  
Db 301 AAGCCTCTGATATGAAGAAACAGAGCTTCTTAAAGATACTCAAAAGAGATGATGCA 360  
QY 409 ACCAAGATCTGACTGAGGAGTAGAGTAGAGATAAGATGCGATTGAAAAATTTGAGC 468  
Db 361 ACCAAGATCTGACTGAGGAGTAGAGTAGAGATAAGATGCGATTGAAAAATTTGAGC 420  
QY 469 AAGAGGACTGTAAAGGTCACAAAGCAACACAAAGAGATTGTAAACGGCTATTAAAGACTT 528  
Db 421 AAGAGGACTGTAAAGGTCACAAAGCAACACAAAGAGATTGTAAACGGCTATTAAAGACTT 480  
QY 529 ATGGGGTTCCTGTTGTAGAGGACCTTCTCAAGCAGACAGATGTCAGCCCTTTGC 588  
Db 481 ATGGGGTTCCTGTTGTAGAGGACCTTCTCAAGCAGACAGATGTCAGCCCTTTGC 540  
QY 589 ATAAACGATAAGGTGTTGCTGTTTCAGAGATATGGACTCCCTTACTTTTGGGGCT 648  
Db 541 ATAAACGATAAGGTGTTGCTGTTTCAGAGATATGGACTCCCTTACTTTTGGGGCT 600  
QY 649 CCAGGTTCTCTGCTCATTTAATGGATCCAAAGTTTCCAAAGAAATACCTGTGATGGAATTT 708  
Db 601 CCAGGTTCTCTGCTCATTTAATGGATCCAAAGTTTCCAAAGAAATACCTGTGATGGAATTT 660  
QY 709 GATGTTGCCAAGGTTTGGAGGAGCTTGAACCTCACCATGACAGTTTCATTTGTTGTCG 768  
Db 661 GATGTTGCCAAGGTTTGGAGGAGCTTGAACCTCACCATGACAGTTTCATTTGTTGTCG 720  
QY 769 ATCCTGTGTGATGTGACTATTGTATAGATCAAAAGTATCGGGGGCAACAGCTCTG 828  
Db 721 ATCCTGTGTGATGTGACTATTGTATAGCATCAAAAGTATCGGGGGCAACAGCTCTG 780  
QY 829 AACTTATTGCTCAACATGGTCCATAGAAAGCATCTTGAGAAATCTTAAATAAGACAGA 888  
Db 781 AACTTATTGCTCAACATGGTCCATAGAAAGCATCTTGAGAAATCTTAAATAAGACAGA 840  
QY 889 TATCAAAATTCCTGAGGACTGGCCCTTACCAGAAAGCTTCGAGCTTGTTCAGAGGACCTAAT 948  
|||||



```
Db 841 TATCAAAATTCCTGAGACTGGCCTTACCAAGAAGCTCGAGCGTTGTTCAAGGAGCCTAAT 900
QY 949 GTCACATTTGATATTCCTGAGCTAAATGAGCTGCACCTGATGAGGAGGTCTCATAAGT 1008
Db 901 GTCACATTTGATATTCCTGAGCTAAATGAGCTGCACCTGATGAGGAGGTCTCATAAGT 960
QY 1009 TTCTCTGTTAAAGATATGTTTCAACGAGAGATCGGTGACAAAGCCATAGAGAAGATC 1068
Db 961 TTCTCTGTTAAAGATATGTTTCAACGAGAGATCGGTGACAAAGCCATAGAGAAGATC 1020
QY 1069 AAATCTGCCAAGATAAATCGTCCGAGGAAGACTCGAGTCCCTTTTCAAGCCAACTGCC 1128
Db 1021 AAATCTGCCAAGATAAATCGTCCGAGGAAGACTCGAGTCCCTTTTCAAGCCAACTGCC 1080
QY 1129 ACCACATCAGCACCGCTAAACCGAAGGAGACTTCGGATATAAACAAGCAAGGAGCTGGC 1188
Db 1081 ACCACATCAGCACCGCTAAACCGAAGGAGACTTCGGATATAAACAAGCAAGGAGCTGGC 1140
QY 1189 AACAGAAGAAACAAGCTGTGGAAGAGAAATAATCTTGGATGCTTGATGTACAACTA 1248
Db 1141 AACAGAAGAAACAAGCTGTGGAAGAGAAATAATCTTGGATGCTTGATGTACAACTA 1200
QY 1249 CGACTACGAAGCAGCGGTGGCTGATCACTTCGCTTAGATATTATTAACCTCCCTGTTTA 1308
Db 1201 CGACTACGAAGCAGCGGTGGCTGATCACTTCGCTTAGATATTATTAACCTCCCTGTTTA 1260
QY 1309 ACTCAGAGCTTTGGTAAAGTTTGCTCATGTTTCAAGCTGGGTAAAGTTAGTTGTTGTTG 1368
Db 1261 ACTCAGAGCTTTGGTAAAGTTTGCTCATGTTTCAAGCTGGGTAAAGTTAGTTGTTGTTG 1320
QY 1369 AAGAGATTGGTGTACCAAGTAAACAAAACCTTATGCTCTGTTTTT 1411
Db 1321 AAGAGATTGGTGTACCAAGTAAACAAAACCTTATGCTCTGTTTTT 1363

RESULT 5
AB021666 1354 bp mRNA linear PLN 19-JUL-2002
LOCUS Oriza sativa (japonica cultivar-group) OSFEN-la mRNA for FEN-1,
DEFINITION complete cds.
ACCESSION AB021666
VERSION AB021666.2 GI:4587224
KEYWORDS Oriza sativa (japonica cultivar-group) (cultivar:Nipponbare)
SOURCE meristem cDNA to mRNA.
ORGANISM Oriza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS Kimura,S., Ueda,T., Hatanaka,M., Takenouchi,M., Hashimoto,J. and
Sakaguchi,K.
TITLE Plant homologue of flap endonuclease-1: molecular cloning,
characterization, and evidence of expression in meristematic
tissues
JOURNAL Plant Mol. Biol. 42 (3), 415-427 (2000)
MEDLINE 20256470
REFERENCE 2 (bases 1 to 1354)
AUTHORS Kimura,S., Hashimoto,J. and Sakaguchi,K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1998) Seisuke Kimura, Science University of
Tokyo, Dept. of Applied Biological Science; 2641 Yamazaki, Noda,
Chiba 278-8510, Japan (E-mail:seisuke@s.noda.tus.ac.jp,
Tel:81-471-24-1501(ex.3409), Fax:81-471-23-9767)
COMMENT On Apr 17, 1999 this sequence version replaced gi:4062866.
FEATURES
Location/Qualifiers
1..1354
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/tissue_type="meristem"
1..1354
/gene="OSFEN-la"

gene
```

```
CDS
67..1209
/gene="OSFEN-la"
/function="endonuclease"
/codon_start=1
/product="FEN-1"
/protein_id="BAA36171.1"
/db_xref="GI:4587225"
/translation="MGIKGLTKLLADNAPKAMKEQKESVFGRRIVADSMSTYQPLI
VYVPTGMEITLNEAGEVTSHQENRNTIRLLEAGIKPVYVDFDKPDLKQELAKRY
SKREDATKELTEAVEEGDKDAIEKFSKRVKVTQKHNECKRLRLMLGVVVEAPCEA
EAECAALCINDMYYAVASDMDSLTGAPRLRHLMDSPSKKI:PVMEFVAVKYLEELE
LTMDFQIDLCISGDCDYSIKIGGQTAALKIIRHGSIESILEINKROYIPEIDMP
YDRAERLFPENVTLDIPELKNWAPDEGLVFLVKEGFMNODRVTKAIEKIKFAPKNK
SSQRLSEFFKPVWTSVPLKRKDTSEKPTKAVANKTKGAGGKKF"
polyA_site
1354
/gene="OSFEN-la"
/notes="17 A nucleotides"
BASE COUNT 415 a 286 c 331 g 322 t
ORIGIN
Query Match 63.4%; Score 927.4; DB 8; Length 1354;
Best Local Similarity 86.3%; Pred. No. 3.4e-221;
Matches 1037; Conservative 0; Mismatches 161; Indels 3; Gaps 1;
QY 55 CCGCCACCCCGCCAGCGCCGAGAGATGGGCATCAAGGGTTTGACGAAACTGCTG 114
Db 37 CGCCGCTCGCGCATCCGTCGGAGCGAGATGGGCATCAAGGGTTTGACGAACTTTG 96
QY 115 GCGGACAATGCGCCCAAGGGGATGAAGGAGCAGAAGTTCGAGAGCTACTTCGGCCGCAA 174
Db 97 GCGGACAATGCGCCCAAGGGGATGAAGGAGCAGAAGTTCGAGAGCTACTTCGGCCGCGC 156
QY 175 ATCCCGCTGCAGCGCAGCATGAGCATATACCAGTTCCTGATTTAGTTGGAAGGACAGGC 234
Db 157 ATTCCTGTCAGCGCAGCATGAGCATCTCCAGTTCCTCATTTGATTTGGAAGGACGGG 216
QY 235 ATGGAACCTCTCACAATGAAGCTGGTGAAGTCACTAGTCACTTTCGAAGGAATGTTCAAC 294
Db 217 ATGGAACCTCTCAGAACGAGCGCGGTGAAGTCACTGATTTGCAAGGTATGTTCAAC 276
QY 295 CGGACAATAGATTACTTGAAGCGGGGAATCAAGCCAGTTTATGTTTGAAGCTTAAAGCT 354
Db 277 CGGACAATAGATTGCTGGAGCGAGGAATCAAGCCAGTATATGTTTGAAGCGAAGCT 336
QY 355 CCTCATATGAAGAAACAAGAGCTTGTCTAAAGATACCTCAAAAAGAGATGATGCAACCAA 414
Db 337 CCTCATGAAGAAACAAGAGCTGCAAAAAGAGTACTCGAAGAGGAGGATGCCACCAA 396
QY 415 GATCTGACTGAGGCGAGTAGAGTAGAGATAAAGATGCGATTGAAAATTCGAGCAAGAGG 474
Db 397 GAACCTAACAGAGGCGAGTAGAGGAAGGAGATAAAGACGCAATTTGAAAATTCAGCAAGA 456
QY 475 ACTGTAAGGTCACAGGCAACACACCAAGATGTAACCGCTATTAAAGACTTATGCGG 534
Db 457 ACTGTGAAGGTCACAGGCAACCAATGAAGAATGTAACCACTATTAAAGACTTATGCGG 516
QY 535 GTTCCTGTTGTAGAGGCACTTCTGAAGCAGAAGCAGAATGTGACCGCTTTTGCATAAAC 594
Db 517 GTTCCTGTTGTAGAGGCGCCATGTGAACGAGCAGAATGTGCTGCATTTTGCATAAAT 576
QY 595 GATAAGGTTTCGCTGTTGCTTTCAGAGATATGAGTCCCTTACTTTTGGGCGTCCAGG 654
Db 577 GATATGTTTATGCAAGTTCATCAGAGATATGGAAGTTCCTTACTTTTGGAGCTCCAGG 636
QY 655 TTCCTTCCTCATTTAATGATCCAGTTTCCAAAGAAAATACCTGTGATGAA-TTTTATGTT 714
Db 637 TTCCTTCCTCATTTGATGGATCCAAAGTTTCCAAAGAAAATTCAGATTTGGAATGGAATC 696
QY 715 GCCAAGGTTTGGAGGAGCTTGAACCTCACCAGTCCAGTTCATTGATTTTCGATCCTG 774
Db 697 GCGAAAGTTTGGAGGAGCTTGAACCTCAGTCTACTATGACCAATTCATTGCTGTCATCCTA 756
QY 775 TGTGATGTGACTATTGTGTAGTACATCAAGGATTCGGGGGGGCAACAGCTCTGAACTT 834
```

```
Db 757 TCTGGATGTCATATTTGTGATAGCATTAAGGTTATGGAGTCAACAGCTCTGAAACTT 816
QY 835 ATTTCGTCAACATGGGTCCTAGAACGATCTTTGGAGAACTTTAATAAAGACAGATATCAA 894
Db 817 ATTTCGTCAACATGGGTCCTAGAACGATCTTTGGAGAACTTTAATAAAGACAGATATCAA 876
QY 895 ATTTCGTAGAGTGGCTTTACCAAGAAGCTCGACGCTTTGTTCAAGGAGCTTAATGTACACA 954
Db 877 ATTTCGTAGAGTGGCTTTATCAAGAAGCTCGACGCTTTGTTCAAGGAACTTCAATGTTACA 936
QY 955 TTGGATATTCCTGACGTAATAGGCTGACCTGATGAGGAGGCTCTCATAGTTTCTGTG 1014
Db 937 TTGGATATTCCTGAGCTGAATAGGATGCTCCGATGAGGAAGGCTTTGTAGAGTTTCTG 996
QY 1015 GTAAAAAATGTTTCAACGAAGCTCGGTGCACAAAGGCGCATAGAGAAGATCAAACTCT 1074
Db 997 GTCAAGAAATGTTTCAACCAAGATAGAGTGACAAAGGCGCATAGAGAAGATTAATTT 1056
QY 1075 GCCAAGAAATGCTGCGCAAGAGACTCGAGTCTCTTTTCAAGCCCACTGCCACCACA 1134
Db 1057 GCCAAGAAATGCTTCCCAAGGAAGGCTCGAATCTCTTCAAGCCAGTTGTTAGCACA 1116
QY 1135 TCAGCACCTTAAACGGGAGAGACTTCGGATAAACAAGCAAGCGAGCTGCCGAACAG 1194
Db 1117 TCAGTCTCTGAAGAAGAGACACTTCAGAAACCAACCAAGCAAGCTCGCTTAACAG 1176
QY 1195 AAAACAAG--GCTGGTGAAGAAGAAATAATCTTGATGCTTGTATCAACTACGA 1251
Db 1177 AAAACAAGGAGCGCGGCGAAGAAATAAGTTGATGCTGATTCATCAACAGA 1236
QY 1252 C 1252
Db 1237 C 1237

RESULT 6
XLU68141
LOCUS
DEFINITION
Xenopus laevis xFEN1b mRNA, complete cds.
ACCESSION
U68141
VERSION
U68141.1 GI:1549392
KEYWORDS
Xenopus laevis.
SOURCE
Xenopus laevis.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 1469)
Bibikova, M., Wu, B., Chi, E., Kim, K. H., Trautman, J. K. and Carroll, D.
Characterization of FEN-1 from Xenopus laevis. cDNA cloning and
role in DNA metabolism
J. Biol. Chem. 273 (51), 34222-34229 (1998)
MEDLINE
99069415
PUBMED
9852084
REFERENCE
2 (bases 1 to 1469)
Bibikova, M., Chi, E., Wu, B., Kim, K. H. and Carroll, D.
Direct Submission
AUTHORS
Submitted (27-AUG-1996) Biochemistry, Univ. of Utah, 50 N. Medical
Drive, Salt Lake City, UT 84132, USA
JOURNAL
Location/Qualifiers
FEATURES
1. 1469
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/tissue_type="head"
/clone_lib="Hemmati-Briuanlou et al. (1991) Development
111, 715-724"
/dev_stage="tadpole"
/notes="also present in oocytes, as determined by
sequencing of RT-PCR products"
120..1268
/function="5'-3' exonuclease and a structure-specific
endonuclease that removes 5' single-stranded 'flaps';
```

```
implicated in processing of Okazaki fragments during DNA
replication and in base excision repair; 5' nuclease"
/notes="FEN-1 homolog; similar to mammalian Dnase IV or
MFL, to S. cerevisiae RAD27, RTH1 or YKL510, and to S.
pombe rad2; similar to Xenopus laevis xFEN1a encoded by
GenBank Accession Number U64563"
/codon_start=1
/product="xFEN1b"
/protein_id="AA08478.1"
/db_xref="GI:1549393"
/translation="MGIHLAKLIADVAPAIKEHDIKSYFGRKVVAVDASMCYIOFLI
AVRDGNMLQNEEGETTSHLMGFKYIRMLEHGKIPVYVFDGPKQKSGSLAKRSE
RAAEAEKLEAEAGEVENIEFKNRLVKTQKNECKKLLSLMGIPYVAPCEAE
ATCAALVKAQVYAAATDEMTALTCTPVLRLHASEAKKLPIDFHLNRYVDIGI
NHEQFVDCILLGSDYCEITRIGIPKRALDLIRQHTLEIINDIDLLKYPPIPEWLH
KEAROLFLEPEVIDADITELKWTEDDEGLVAFMCGEKQFSEIRIRNGAKKLAKNRQG
STQGRIDDFEFKVTGSISSYTKRKEVSKSTKKSKTKTGGTPAGFKRKG"
BASE COUNT 461 a 288 c 377 g 343 t
ORIGIN
Query Match 22.3%; Score 326.8; DB 5; Length 1469;
Best Local Similarity 58.1%; Pred. No. 8.5e-71;
Matches 615; Conservative 0; Mismatches 437; Indels 6; Gaps 2;
QY 78 AGACGAGATGGCATCAAGGGTTTGACGAACTGCTGGCGGACAATCGGCCCAAGCGAT 137
Db 113 AGTCAATATGGAAATTCACGGCTTGCCAAAGCTTATCGCAGATGTGCGCCTGCACCTAT 172
QY 138 GAAGGACGAGAAGTTCGAGAGCTACTTCGGCGCAAAATCGCGTCGCGCAGCATGAG 197
Db 173 CAAAGAGCATGATATAAAAGTTACTTTGGTCGAAAAGTGGCGGTTGATGCTTCTATGTG 232
QY 198 CATATACCACTTCTGATTTGATGTTGGAAGGACAGGACGATGGAACCTCTCACAAATGAAGC 257
Db 233 CATTTACCAGTTCTTATCGCTGTGACAGAGGATGSCA---ATATGCTGCAGAAATGAGGA 289
QY 258 TGTGTAAGTCACTAGTCAATTTGCAAGAAATGTTCAACCGGACAATAAGATTACTTGAAGC 317
Db 290 AGCGAAACACCAAGCATCTCATGGGTATGTTTACCCTAGCAATTCGCAATGTTGAGCA 349
QY 318 GGAATCAAGCGAGTTTATGTTTTCGTCGAGCTCTCTGATATGAACAAACAGAGCT 377
Db 350 TGGCATCAAAACCCGTTATGTTTGTGGAGGCGCCCTCAGATGAAGTCAAGTGAAGT 409
QY 378 TGCTAAAGATCTCAAAAAGAGATGATGCAACCAAGATCTGACTGAGCGAGTAGAGT 437
Db 410 AGCCAAAGCGAGTGAGCGTAGGCGACAGGAGAGTTGCTTGAAGCTGCCGAGGAGCG 469
QY 438 AGAGATAAAGATGCGATTCAGAAATTTGACGAGAGGAGCTGTAAGGTCACAGGCAACA 497
Db 470 TGGAGAAGTGGAGAACATTTGAGAAGTTCAATAAAAGGCTTGTCAAGGTCAACCAAGCAGCA 529
QY 498 CAACGAAGATTGTAACCGCTATTAGACTTATGGGGTTCTCTGTTAGAGGACACCTTC 557
Db 530 CATGAAGAGTGCAGAAATTCGTTAGCCTGATGGGTATTCCTTATGTTGATGTCACCTG 589
QY 558 TGAAGCAGAGCAGAAATGTGACGCCCTTTGCATAACGATAAGGTTCTGCTGTTGCTTC 617
Db 590 TGAAGCTGAGGCCACATGCGCTGCTTAGTAAAGCAGGAAAAGTGTATGCTGCGAGCAAC 649
QY 618 AGAAGATATGGACTCCCTTACHTTTTGGGCTCCAGGTTCTCTGCTCATTTAATGGATCC 677
Db 650 TGAAGACATGGATGGCTTACCTTTGGCACTCCGTTGCTTAGGCACTCACAGCAAG 709
QY 678 AAGTTCCAAGAAATACCTGTGATGAAATTTGATGTTGCCAAGGTTTGGAGGAGCTTGA 737
Db 710 TGAAGCTAAAGCTGCTTATTCAGAGTTCCTTAACCCGCTTTTCAGGACATCGG 769
QY 738 ACTCACCATTGGACCAAGTTCATTTGATTTGTCATCTGTTGGATGTGACTATTTGATAG 797
Db 770 CATCAATCAGAACAAATTTGTGATCTCTGTATCTTGTGCGAAGTGAATTTATTTGAAC 829
QY 798 CATCAAGGTATCGGGGGGCAACAGCTCTGAAACTTATTCGTCAACATGGGTCCATAGA 857
```



```
Db 1128 CTGGATGACTTTTCAAAGTACCGGATCCGTTAGCTCAACTAAAGAAAGAGGCGAGAA 1187
QY 1162 TCGGATAAACAAGCAGCGCTCGACACAGAAACAAGGCTGGTGGAAAGAGAA 1220
Db 1188 TCCAAAGGATCCGCTAAAGAAAGACCAAACTGGAGGTACCCCTGCGAGGAAGTTCAA 1246

RESULT 8
AF065397 1357 bp mRNA linear VRT 05-JAN-1999
LOCUS xenopus laevis flap endonuclease 1 (FEN1) mRNA, complete cds.
DEFINITION AF065397
ACCESSION AF065397
VERSION AF065397.1 GI:4106357
KEYWORDS
SOURCE Xenopus laevis.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE Li, J.-L. and Cox, L.S.
AUTHORS Cloning and investigation of Xenopus FEN1: developmental expression
TITLE and function in DNA replication
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1357)
Li, J.-L. and Cox, L.S.
DIRECT SUBMISSION
TITLE Submitted (14-MAY-1998) Department of Biochemistry, University of
Oxford, South Parks Rd., Oxford OX1 3QU, England
JOURNAL Location/Qualifiers
FEATURES
source 1..1357
/organism="Xenopus laevis"
/db_xref="taxon:8355"
1..1357
/gene="FEN1"
99..1247
/note="Involved in DNA replication and repair"
/codon_start=1
/product="flap endonuclease 1"
/protein_id="AAD02814.1"
/db_xref="GI:4106358"
/translation="MGIRGLAKLIADVAPAIKEHDIKSYFGRKVAVDASMCYQFLI
AVROGNTLQNEGETTSHLMGFYTRIMVEHGKIPYVDFGRPKMSSELAKRSE
RAAEKLEAEAEAGEVENIEKTRKLVKTKOHNSECKLLTLMLGIPYVDAPCEAE
ATCAALVRAGVAAATDMDATFTGLVLLHLTASEAKLPLOEFLNVRVODIGI
THEOFDLICLLGSDYCEITRIGIGPKRIDLIRHKTIDEIIDNIDLKYPVPENLH
KEAKHLFLEPVEDVDITELKWIPEDEGLVAFMGKOFSEDRIRNGAKKLAKNRQG
STQGRLLDFFKVTGVSYSSTKREAEKSGSAKKKAKTGTTPAGKFKRGR"
BASE COUNT 425 a 284 c 345 g 303 t
ORIGIN
Query Match 21.0%; Score 307; DB 5; Length 1357;
Best Local Similarity 55.6%; Pred. No. 7.6e-66;
Matches 633; Conservative 0; Mismatches 500; Indels 6; Gaps 2;
QY 85 ATGGGCATCAAGGTTTTCAGCAACTGCTGGCGGACAAATGCGCCCAAGGGATGAAGAG 144
Db 99 ATGGGAATTCACGGTTTGGCCAAACTTATCGCAGATGTGGCACCTGCAGCATATCAAGAG 158
QY 145 CAGAAGTTCGAGACTACTTTCGGCCGCAAAATCGCGTCGACCGCAGCATGAGCATATAC 204
Db 159 CATGATATAAAAGTTACTTTGGTCGAAAGGTGGCGGTGCGATGCTTCATGTCATTTAT 218
QY 205 CAGTTCCTGATTGTAGTTGGAAGACAGCAGGATGGAACCTCTCAAAATGAAGTGGTGA 264
Db 219 CAGTTTCTATTCAGTGAGACAGGATGGCA---ACACGCTGCAAAACGAGGAAGTGAA 275
QY 265 GTCATAGTCAATTTGCAAGGAATGTTCAACCGGACAAATAGATTACTGGAACGGGAATC 324
Db 276 ACCACAAGCCATCTATGGGCATGTTTACCCTGACCATGCGCATGGTAGAGATGGCATC 335
QY 325 AAGCCAGTTTATGTTTTTGTGGCAAGCCCTCTGATATGAAGAAACAAAGAGCTTGTGTA 384
```

```
Db 336 AAACCAAGTGTATGTGTTGATGGAAAGCCACACAGATGAAGTCCGCTGAACCTAGCCAAA 395
QY 385 AGATACTCAAAAAGAGATGATCAACCAAGATCTCACTGAGCGAGTAGAGGTAGAGAT 444
Db 396 CGAGTGAAGCTAGGCGGAGGAGAGAGTTCCTTGAAGCTGCCGAGGAGGTGGAGAA 455
QY 445 AAAGATGCGATTGAAAATTTGACCAAGAGAGACTGTAAAGGTCAACAGGCAACACACGAA 504
Db 456 GTGAGAACATTTGAGAAGTTTCACTAAAGACTTGTCAAGGTTTACCAAGCAGCACAATAG 515
QY 505 GATTGTAACCGGTTTAAAGACTTATGGGGGTTCTCTGTGTAGAGGCACTTCTGAAGCA 564
Db 516 GAATGCAAGAAATTAACCTACCCCTAATGGGCAATTCCTTATGTGGTGCACCTGTGAAGCT 575
QY 565 GAAGCAGAATGTGAGCGCTTTGCATAAAGATGAAGGTTCCTGCTTGTGCTTCAAGAGAT 624
Db 576 GAGCCACAGTGTGCTGCTTAAAGAGCAGGAAAGTGTATGCTGCAGCAACAGAAGAC 635
QY 625 ATGGACTCCCTTACTTTTGGGGCTCCACGGTTCCTCTGCTCATTTAATGGATCCAAGTTC 684
Db 636 ATGGATCCCTGACCTTTGGCACCCCTGTCTCTTAGGCACCTCACAGCAAGCGAAGCT 695
QY 685 AAGAAATACCTGTGATGGAATTGATGTTGCCAAGGTTTGGAGGAGCTTGAACCTCAC 744
Db 696 AAAAAAGCTGCTATTCAGGAGTTCCATCTAAACCGTGTCTATTCAGGACATAGTATCACA 755
QY 745 ATGGACCAAGTTTCAATTGATGTCATCTGTGGATGTGACTTATTTGTATGATACATCAA 804
Db 756 CATCAACAAATTTGGGATCTCTGTATCTTCTCGGAAGTGACTACTGTGAAGACATCCGT 815
QY 805 GGTATCGGGGGGAAACAGCTCTGAAACTTATTCGTCACATGGTTCCTATAGAAAGCATC 864
Db 816 GGCATTGGACCAAAAGAGCCATTGACCTATCGCTCAGCATTAAGCATAGACATTGAGGATC 875
QY 865 TTGGAGAAATCTTAATAAAGACAGATATCAAAATCTCTGAGGACTGGCTTACCAGAGCT 924
Db 876 ATTGATAACATTGACCTCAAGAAATACCCCTTCCGAAAACCTGGCTGCACAGGAGGCA 935
QY 925 CGACGCTTTGTTCAAGGAGCCCTAA---TGTCACATTTGATATTTCTGAGCTAAATGACT 981
Db 936 AAACATCTTTTCTTGGAACCAAGAGTAGTCACACTGACATCACCGAGCTGAAGTGGATC 995
QY 982 GCACCTGATGAGGAGGCTTCATAAGTTTCTGCTGTAAGAAATATGTTTCAACGAAGAT 1041
Db 996 GAGCCCGATGAAGAGAGATTGGTGGCTTCTATGTTGGGAGAGAGCTTTCAGTGAAGAC 1055
QY 1042 CGGTTGCAAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAAATAAATCGTCCAAAGGA 1101
Db 1056 CGCATACGCAACGGGGGCCAAAAGTTGGCCAAAGAACCGGCAAGGCACACACAGGGCGG 1115
QY 1102 CTCGAGTCTTTTTCAGCCCAACTGCCACCATCAGCAGCGCTTAAACGGAAGGAGACT 1161
Db 1116 CTGGATGACTTTTTCAAAAGTACCGGATTCCTGTAGCTCACTAAAGAAAGAGGCGAGAA 1175
QY 1162 TCGGATAAAACAAAGCAGGAGCTCGCAACAAAGAAACAAAGGCTGGTGGAAAGAGAA 1220
Db 1176 TCCAAGGATCCGCTTAAAGAAAGCCAAACTGGAGGTACCCCTGCGAGCAAGGTTCAA 1234

RESULT 9
AF036327 1149 bp mRNA linear VRT 26-OCT-2000
LOCUS Xenopus laevis flap endonuclease 1 (FEN1) mRNA, complete cds.
DEFINITION AF036327
ACCESSION AF036327
VERSION AF036327.1 GI:2674206
KEYWORDS
SOURCE Xenopus laevis.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 1149)
```

AUTHORS Kim,K., Biade,S. and Matsumoto,Y.  
TITLE Involvement of flap endonuclease 1 in base excision DNA repair  
J. Biol. Chem. 273 (15), 8842-8848 (1998)  
MEDLINE 98204872  
PUBMED 9535864  
REFERENCE 2 (bases 1 to 1149)  
AUTHORS Kim,K. and Matsumoto,Y.  
TITLE Direct Submission  
JOURNAL Submitted (01-DEC-1997) Radiation Oncology, Fox Chase Cancer  
Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA  
FEATURES  
Location/Qualifiers  
1..1149  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/cell\_type="oocyte"  
1..1149  
/gene="FEN1"  
1..1149  
/gene="FEN1"  
/function="structure-specific nuclease"  
/codon\_start=1  
/product="flap endonuclease 1"  
/protein\_id="AAB88707.1"  
/db\_xref="GI:2674207"  
/translation="MGHGLAKLIADVAPAAIKHEIDIKSYFGRKVAIDAMSGIYQFLI  
AVROGNLQNEGETTSHLMGFYRTIRMVEHGKIPVYFDGKPKOMKSGELAKRSE  
RRAEAKLEAAEEAGEVENIEKFKRLVKVTKQHNBECKLLTILMGIPYVDPCEAE  
ATCAALVAKAGKYAAATEMDALFTGTPVLLRLHTASEAKKLPIQEFHLNRYQDIGI  
THEQVDFCILLGSYDCEITRIGIKPRAIDLRQHKTDIEIDNIDLKYPVPEWNLH  
KEAKHLFDEPVDVDITELKWIPEDEGLVAFMGCEKQFSEDRIRNKAKLAKNRQG  
STQGRLLDFFKVTGVSSTKRKEAKSGSAKKKAKTKGTGTPAGKFRKG"  
BASE COUNT 352 a 247 c 306 g 244 t  
ORIGIN  
Query Match 20.9%; Score 305.4; DB 5; Length 1149;  
Best Local Similarity 55.5%; Pred. No. 1.8e-65;  
Matches 632; Conservative 0; Mismatches 501; Indels 6; Gaps 2;  
QY 85 ATGGGATCAAGGGTTGAGCAACTCTCGCGGACAATCGCCCAAGGCGATGAAGAG 144  
DB 1 ATGGGAATTCACGGTTGGCCAACTATTCGCAGATGTGCACCTGCAGCTATCAAGAG 60  
QY 145 CAGAAGTTCAGAGCTACTTCTCGCGCAAAATCGCCGTCGACGCGACGATGAGCATATAC 204  
DB 61 CATGATATAAAGTTACTTTGGTGAAGAGTGGCGATCGATGCTTCCATGTGCAATTTAT 120  
QY 205 CAGTCTCGATTTAGTTGGAAGACAGGATGGAAGACTCTCAAAATGAAGCTGGTGA 264  
DB 121 CAGTTCTTATTTGGGTGAGACAGAGTGGCA---ACACGCTGCAAAACGAAGAGGTGA 177  
QY 265 GTCACTAGTCAATTCGAAGGAATGTTCAACGGACAATAAGATTACTGGAAGCGGAATC 324  
DB 178 ACCAAGCCATCTATGGGTATGTTTACGTTACCATTCGATGCTAGGACGATGCAATC 237  
QY 325 AAGCCAGTTTATGTTTGTGGCAAGCTCTCTGATATGAGAACAAGAGCTTGCTAAA 384  
DB 238 AAACCAAGTATGTTTGTGGGAAGCGCCACAGATGAAGTCGGGTGAACAGCAAA 297  
QY 385 AGATCTCAAAAGAGATGATGCAACCAAGATCTGACTAGGACAGTAGAGTAGAGAT 444  
DB 298 CGCAGTGAGCGTAGGCGGAGGCAAGAGTTGCTTGAAGCTGCGGAGGCGGTGGAGAA 357  
QY 445 AAAGTGGATTTGAAATTCAGCAAGAGGACTTAAGGTCAACAGGCAACACAAACGAA 504  
DB 358 GTGGAGAACATTTGAGAAGTTTCACTAAGAGACTTGTCAAGGTTTCAAGCAGCACAAATGAG 417  
QY 505 GATTGTTAAACGGCTATTAAAGACTTATGGGGTTCCTGTTGTAGAGGACCTTCTGAAGCA 564  
DB 418 GAGTGCAGAGAATTACTTACCTTAATGGCATTCCTTATGTGGATGCCACCGTGTGAAGCT 477  
QY 565 GAAGCAGAAATGTGACGCCCTTTGCATAAACGATAGGTGTTGCTGTTCAGAGAT 624  
DB 478 GAGGCCAGTGTGCTGCCTTAGTAAAGCAGGGAAGTGTATGCTGCAGCAACAGAGAC 537

QY 625 ATGGACTCCCTTACTTTTGGGCTCCACGGTTCCTTCGTCATTATATGATCCAAGTTCC 684  
DB 538 ATGGATGCCCTGACCTTTGGCAGCCCTGTGCTCCTTAGCAGCTACACAGACGAGCT 597  
QY 685 AAGAAAATACCTGTGTATGGAATTTGATGTTGCCAAGGTTTGGAGGAGCTTGAACAC 744  
DB 598 AAAAGCTGCGCTATTACAGGAGTTCCATCTAAACCGTGTTCATTACAGGACATAGTAAC 657  
QY 745 ATGGACCACTTCAATGATTTGTCATCCTGTGTTGATGTTGATGTTGATGATGATGAT 804  
DB 658 CATGAACAATTTGCGATCTGTATCTTCTGCGGAAGTGACTACTGTGAACCATCCGT 717  
QY 805 GGTATCGGGGGCAACAGCTCTGAAACCTTATTCGTCAACATGGTCCATAGAAAGCATC 864  
DB 718 GGCATTGGACCAAAAGAGCCATTGACCTGATCCGTCAGCATAAGACCATTCACGAGATC 777  
QY 865 TTGGAGAATCTTAATAAGACAGATATCAAAATTCCTTGAGGAGTGGCTTACCAGAGCT 924  
DB 778 ATTGATAACATTGACCTCAAGAAATACCCCGTTCCCGGAAACATGGCTGCAAGAGGCA 837  
QY 925 CGAGCGCTTGTCAAGGAGCTAA---TGTCACATTGGATATTCCTGAGCTATAAATGACT 981  
DB 838 AACATCTTTCTTGGAACCAAGAGTAGTAGTCACACTGACATCACCAGCTGAAGTGATC 897  
QY 982 GCACCTGATGAGGAGGCTCTCATTAAGTTTCTGTTAAAGATAATGTTTCAACAGAGAT 1041  
DB 898 GAGCCCGATGAAGAAGATTGGTGGCTTCATGTGGGGAAGAGCAGTTTGTAGTGAAGAT 957  
QY 1042 CGGGTGACAAGGCCATAGAGAAGATCAATCTGCCAAGATAAATCGTGCAGAGGAAGA 1101  
DB 958 CGCATACGCAACGGGGCCAAAAAGTTGGCCAAAGACCGGCAAGCAGCACACAGGCGCG 1017  
QY 1102 CTGAGTCTTTTTCRAAGCCAACTGCCACCATCATCAGCAGCGCTAAACAGGAGGACT 1161  
DB 1018 CTGATGACATTTTCAAGTGACCGGATCGTTCAGTCACTAAAGAAAGAGGAGAA 1077  
QY 1162 TCGGATAAAACAAGCAGGAGCTGCGAACAAGAAACAAGGCTGTTGGAAGAGAA 1220  
DB 1078 TCCAAAGGATCCCGTAAAGAAAGCCAAACTGAGGTACCCCTGCGAGGAAGTTCAA 1136  
RESULT 10  
BC010203 2163 bp mRNA linear ROD 07-AUG-2002  
LOCUS Mus musculus, Similar to flag structure-specific endonuclease,  
DEFINITION clone MGC:7137 IMAGE:3158176, mRNA, complete cds.  
ACCESSION BC010203  
VERSION BC010203.1 GI:16307327  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2163)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-JUL-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Center  
Center code: BCM-HGSC  
Web site: <http://www.hqsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 6 Row: d Column: 19.

# FEATURES

source

Location/Qualifiers  
1. 2163  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB/N"  
/clone="WGC:7137 IMAGE:3158176"  
/tissue\_type="mammary tumor. C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."  
/clone\_lib="NCI\_CGAP\_Mam6"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
206. 1348  
/codon\_start=1  
/product="Similar to flag structure-specific endonuclease"  
/protein\_id="AAH10203.1"  
/db\_xref="GI:16307328"  
/translation="MGHGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLI  
AVRGDVLQNEEGTTHLMGMFYRTIRMMENGIKPVYVDFGKPOLKSELAKRSE  
RRAEAKLOQOAEAGMEVEKFKRLVKVTKOHNDECKHLLSLMGIPYIDAFSEAE  
ASCAALAKAGVYAATDEMCLFSGPVMRLHLSASEAKKLPIOEHLKSLVQELGL  
NOEYFDLILGSDYCSIRIGIKPRAVDLIQKHLSIEEIVRLDPSKYPVPEWHL  
KQAQLFLEPEVLDPESELKWESEELKFMCKEKFSEERIRSGKRLSKRSROG  
STQGRLDLFFKVTGSLSKAKRKEPKPAKKKARTGAGNFRGK"  
540 a 528 c 619 g 476 t

# BASE COUNT

ORIGIN

Query Match 20.9%; Score 305.4; DB 10; Length 2163;  
Best Local Similarity 56.3%; Pred. No. 2.le-65;  
Matches 642; Conservative 0; Mismatches 481; Indels 18; Gaps 3;

QY	85	ATGGGATCAAGGGTTGACGAACTGCTGGCGCAATGCGCCCAAGCGCATGAAGGAG	144
DB	206	ATGGGAATTCACGGCTTCCCAACTAATGCTGATGTGGCCCCAGTGCCATCGGTGAG	265
QY	145	CAGAAGTTCGAGAGTACTTCCGCGCAAAATCGCGTCGAGCGCCAGCATGATATAC	204
DB	266	AATGACATCAAGAGTACTTTGTGCGCAAGTGGCCATCGATCGCTCCATGAGCATATC	325
QY	205	CAGTTCCTGATTGTTGGAAGGACGACATGGAACCTCTCACAAATGAAGCTGGTAA	264
DB	326	CAGTTCCTGATTGCTGCTGAG--GTTGGGATGTGTCGCAAGACGAGGAGGTGAG	382
QY	265	GTCATGATGATTTGCAAGGAATGTTCAACCGGACAATAAGATTACTGGAAGCGGGAATC	324
DB	383	ACCACGACGACCTGATGGCATGTTCTACCGTACCATCCGATGATGAGATGGCATC	442
QY	325	AAGCCAGTTATGTTTGTGATGGCAAGCTCTCTGATGATGAAGAACAAGAGCTTGTAAA	384
DB	443	AAGCTGTGTAGCTTTGTGATGGCAACACACAGCTGGAAGTCAAGTGGCGAGTGGCCAA	502
QY	385	AGATACTCAAAAGAGATGATGCAACCAAGATCTGACTGAGGAGTAGAGTAGAGAT	444
DB	503	CGCAGTGAGAGCGCGCGAGGCTCAGAGCAACTGCGAGAGGCTCAGGAGGCTGGGATG	562
QY	445	AAAGATGCGATTGAATAATTGAGCAAGAGGACTGTAAAGGTGCAAGGACACACAGGAA	504
DB	563	GAGGAGGAGGTGGAGAGTTCCACCAAGAGGCTCGTGAAGGTGCAACCAACACATGAT	622
QY	505	GATTTCAACGCGCTATTAAAGACTTATGGGCTTCTGTTGATAGGACACCTTCTGAAGCA	564
DB	623	GAGTCAACACCTGCTGAGCCTCATGGCATCCCTTACCTTTGATGACCCAGCGAGCA	682
QY	565	GAAGCAAGATGCGCCCTTTTGCATAAACGATAAGGTGTTGCTGTGCTTTCAGAAGAT	624
DB	683	GAGGCGAGCTGTGCTGCTGGCAAGGCTGGCAAGAGTCTATGCTGCGGCCACGAGGAC	742
QY	625	ATGGACTCCCTTACTTTTGGGGCTCCAGGTTCTCTCGTCATTTAATGATGATCAAGTTCC	684

DB	743	ATGGACTGCCTCACTTTTGGCAGCCCCGTGTAATGCGACACTTAATGCCAGCTGAGGCC	802
QY	685	AAGAAAAATACCTGTGATGGAATTTGATGTTGCCAAGGTTTTGGAGGAGCTTGAACCTACC	744
DB	803	AAGAAGCTGCCCATCAAGAGTTCCATCTGAGCCGCGTCTGCGAGGAGCTGGTCTGAC	862
QY	745	ATGGACCACTTCAATGATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	804
DB	863	CAGGAGCAGTTTGTGGATCTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	922
QY	805	GSTATCGGGGGGCAACAGACTTCTGAACTTATTCGTCAACATGGTCCATAGAAAGCATC	864
DB	923	GGCATGGGCCCAAGCGGCTGTGATCTCATCCAGAAACATAAGACATCGAGGAGATC	982
QY	865	TGGGAGATCTTAATAAGACAGATATCAAAATTCCTGAGGAGTGGCCTTACCAGAAAGCT	924
DB	983	GTGAGGGCGCTGGACCCAGCAAGTACCCGCTTCCAGAGAACTGGCTCCACAAGGAAGCC	1042
QY	925	CGAGCTTTTCAAGGAGCCCTAATG---TCACATTGCTATTCCTGAGCTTAAATGGACT	981
DB	1043	CAGCAGCTTCTTCTGGAGCCAGAGTACTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC	1102
QY	982	GCACCTGATGAGGAGGTCTCATTAAGTTTCTGTGTAAGATTAATGTTTCAACGAAGAT	1041
DB	1103	GAGCAAAATGAAGAAGAGTTGGTCAAAATTTATGTGTGTTGTAAGAGAGTTTCTGAAGAG	1162
QY	1042	CGGTGTCAAAAGGCCATAGAGAGATCAAAATCTGCCAAGAAATAATCTGCGCAAGAGA	1101
DB	1163	CGAATTCGAGTGGGTCAAGCGGCTGAGTAAGAGCGCCAGGCGACACCCAGGAGCGC	1222
QY	1102	CTCGAGTCTTTTCAAGCCAACTGCCACCATCTAGCACCCTGCTTAAACGGAAGAGACT	1161
DB	1223	CTCGATGATTCTTCAAGGTGACAGGCTCACTCTCTCTCAGC-----TAAGCGC	1270
QY	1162	TCGATTAACACAGCAGGAGGCTGGCAACAAGAAACAAAGCTGGTGGAAAGAGAAA	1221
DB	1271	AAGGAGCCAGAACCCAGGGGCTGCTAAGAGAAAGCAAGACTGGGGGAGCGGGAAG	1330
QY	1222	T 1222	
DB	1331	T 1331	

RESULT 11  
AC026761

LOCUS  
Mus musculus Strain C57BL6/J chromosome 19 BAC, RP22-325J22,

DEFINITION  
Complete Sequence, complete sequence.

ACCESSION  
AC026761

VERSION  
AC026761.17 GI:22213163

KEYWORDS  
HTG.

SOURCE  
house mouse.

ORGANISM  
Mus musculus

REFERENCE  
1 (bases 1 to 163475)

AUTHORS  
Grills,G., Han,J., Shim,C., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R., Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A., Gordon,M., Goltz,J.S. and Kucherlapati,R.

TITLE  
Mouse High Throughput Sequencing

JOURNAL  
Unpublished

REFERENCE  
2 (bases 1 to 163475)

AUTHORS  
Grills,G., Han,J., Shim,C., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R., Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A., Gordon,M., Goltz,J.S. and Kucherlapati,R.

TITLE  
Direct Submission

JOURNAL  
Submitted (24-MAR-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

REFERENCE  
3 (bases 1 to 163475)

AUTHORS  
Grills,G., Han,J., Shim,C., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R., Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,

Gordon, M., Goltz, J.S. and Kucherlapati, R.  
 Direct Submission  
 Submitted (29-JUN-2002) Harvard Partners Center for Genetics and  
 Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA  
 02139, USA  
 4 (bases 1 to 163475)  
 Grills, G., Han, J., Montgomery, K.T., Lee, E., Long, J., Pomerantz, R.,  
 Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,  
 Gordon, M., Goltz, J.S. and Kucherlapati, R.  
 Direct Submission  
 Submitted (13-AUG-2002) Harvard Partners Center for Genetics and  
 Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA  
 02139, USA  
 On Aug 13, 2002 this sequence version replaced gi:21629248.  
 -----Genome Center:  
 Center: Harvard Partners Genome Center  
 Center Code: HPGC  
 Web site: <http://www.hpcgg.org/Sequence/mouse.html>  
 Contact: hpgc@model.mgh.harvard.edu

CLONE LENGTH: This sequence represents the entire insert of this  
 clone unless otherwise noted. If there are overlapping clones, the  
 overlaps are noted in the beginning and end of the Features  
 listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550).  
 Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.  
 Genes and Regions of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST  
 and cDNA sequences in Unigene. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintain sequence continuity  
 across the splice junctions. Sequences that are not identical  
 matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double  
 strandsequence for all regions. All sequence is completed to a  
 standard of coverage with a minimum of 3 reads with no ambiguities.  
 If the sequence coverage for a region does not meet this standard,  
 it is indicated in the annotation as Low Coverage. Low coverage  
 linkages are verified by PCR product size verification or  
 verification of forward and reverse reads from clones which span  
 the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated average error rate is less than 1 per 10,000  
 bases using the Consed quality parameters. Regions that do not  
 meet this requirement are annotated as Low Quality.

#### -----Summary Statistics

Center project name: ACD  
 Sequencing vector: pUC18, L08752, 36%; pSMART, AF399742, 64%  
 Chemistry: Dye-terminator Big Dye, 100%  
 Assembly program: Phrap version 0.990319  
 Contig length: 163475  
 Fraction of Phrap value < 40: 0  
 Error Rate in Consed: 0.01 per 10,000 bases  
 Number of N's in consensus: 0

----- Distribution of Quality < 40 Bases:

```

1000|
 900|
 800|
 700|
 600|
 500|
 400|
 300|
 200|
 100|
  0|
  
```

FEATURES	source	Location/Qualifiers	Phrap Value Range
repeat_region	repeat_region	1..163475 /organism="Mus musculus" /strain="C57BL6/J" /db_xref="taxon:10090" /chromosome="19" /clone="RP22-325J22" 5605..5715 /rpt_family="PB1D7" complement(6210..6379) /rpt_family="RSINE1" complement(6398..6582) /rpt_family="B4" complement(6583..6732) /rpt_family="BI_MM" complement(6733..6830) /rpt_family="B4" 6849..6868 /rpt_family="ID4" 6869..6933 /rpt_family="ID_B1" 6934..6952 /rpt_family="ID4" 7357..7405 /rpt_family="L2" /product="BC027295 Mus musculus, clone MGC:28169 IMAGE:3985840, mRNA, complete cds" complement(9503..9582) /rpt_family="ID4" 1019830..9862.39925..39927.154245..154609, 15944..156065.156184..156277.156407..156500, 156797..156891.157305..157380.157821..157883, 158140..158262.158368..158470.159437..159651, 160175..160416.160424..160564.160566..160644) /product="AB041560 Mus musculus brain cDNA, clone MNCb-0629, similar to Homo sapiens delta-5 fatty acid desaturase (CYB5RP) mRNA" complement(9848..10039) /rpt_family="B2_Mm1" complement(10211..10351) /rpt_family="BI_MM" complement(10240..10384) /rpt_family="MT2B" 10990..11138 /rpt_family="BI_MM" 12753..12867 /standard_name="AM538437" complement(13788..14159) /rpt_family="MT2A" complement(14471..14534) /rpt_family="B1F" complement(15278..15541) /rpt_family="B4A" join(15583..17616.17873..17998.18090..18181.18407..18494, 18692..18787.19500..19578.20990..21049.25606..25738, 26014..26114.26610..26802.27450..27561.29224..29514) /product="AB072976 Mus musculus D5D mRNA for delta-5 desaturase, complete cds" 20080..20108 /rpt_family="(TCCC)n" complement(20118..20270) /rpt_family="RSINE1" complement(20277..20415) /rpt_family="B4A" 20529..20612 /rpt_family="PB1D9" complement(21436..21636) /rpt_family="B3"	5 10 15 20 25 30 35 40

```
repeat_region 21770..21790
/rpt_family="AT rich"
repeat_region complement(22048..22179)
/rpt_family="L2"
repeat_region 22631..23002
/rpt_family="MSTD"
repeat_region complement(23407..23560)
/rpt_family="B4"
repeat_region 24229..24294
/rpt_family="MER2"
repeat_region complement(24348..24443)
/rpt_family="PB1D10"
repeat_region complement(24492..24568)
/rpt_family="ID4"
repeat_region 24696..24802
/rpt_family="RSINEL"
repeat_region 24895..25051
/rpt_family="RSINEL1"
repeat_region 25053..25166
/rpt_family="MER2"
repeat_region 25764..25927
/rpt_family="B3"
repeat_region 29777..29800
/rpt_family="AT rich"
repeat_region complement(29915..30038)
/rpt_family="ID.B1"
repeat_region 30336..30384

Query Match 20.9%; Score 305.4; DB 10; Length 163475;
Best Local Similarity 56.3%; Pred. No. 4.7e-65;
Matches 642; Conservative 0; Mismatches 481; Indels 18; Gaps 3;

QY 85 ATGGGATCAAGGTTGACAAACTGCTGGGGCAATGCGCCCAAGCGATGAAGAG 144
DB 11409 ATGGGATTCACGCTTGCACAACTAATGCTGTGGCCCGCCAGTCCGTGAG 11468

QY 145 CAGAATGTCAGAGCTACTTGGCGCCGCAAAATCGCGTCGCGCCAGCATGATATAC 204
DB 11469 ATGACATCAAGAGCTACTTGGTCGCAAAAGTGCGCATCGATGCTCCATGAGCTATC 11528

QY 205 CAGTTCCTGATGTAGTGGAGGACAGGATGGAACCTCTCACAAATCAAGCTGGTGA 264
DB 11529 CAGTTCCTGATGTAGTGGTTCGTCAG--GGTGGGATGTGTCGAGAACGAGGAGGTGAG 11585

QY 265 GTCATAGTCAATGCAAGAAATGTTCAACCGACAAATAGATTACTGSAAGCGGGAATC 324
DB 11586 ACCACGACCTGATGGGCAATGTTCTACCGTACCATCCGATGATGGAATGGCATC 11645

QY 325 AAGCCAGTTATGTTTGTATGGCAAGCTCTCTGATATGAAGAAACAGAGCTTGTAA 384
DB 11646 AAGCCTGTGACGCTTTGTATGGCAACACCAACACAGCTGAAGTCAGGCGAGCTGGCCAAG 11705

QY 385 AGATACTCAAAAGAGATGATGAACCAAGATCTGACTGAGGCGAGTAGAGTAGGAGAT 444
DB 11706 CCGAGTGAGGCGCGCGAGGCTGAGAGCAACTCGACAGGCTCAGGAGGCTGGGATG 11765

QY 445 AAGATGCGATTGAAAATTCACAGAGAGCTGTAAAGTTCACAGGCAACACACAA 504
DB 11766 GAGGAGGAGTGGAGAGTTCACAGAGGCTCGTGAAGGTCCACCAAGCAACACATGAT 11825

QY 505 GATTGTAACGGCTATTAAGACTTATGGGGTTCCTTGTGTAGAGGCGACCTTCTGAAGCA 564
DB 11826 GAGTGAACACACCTGCTGAGCCTCATGGGCATCCCTTACCTTGTATGACCCAGGAGGCA 11885

QY 565 GAAGCAGAAATGTCAGCGCTTTCATAAACGATGATGCTGCTGCTTTCAGAAAGAT 624
DB 11886 GAGGCGAGCTGTGCTGCTGCGCAAGGCTGGCAAGTCTATGCTGCGCCAGGAGGAG 11945

QY 625 ATGAGCTCCCTTACTTTTGGGGCTCCACGGTTCCTTGTGTCATATTAATGATCCCAAGTTCC 684
DB 11946 ATGAGCTCCCTTACTTTTGGCGAGCCCGCTGAATGCGACACTTAATGCGAGTGGAGGCC 12005

QY 685 AAGAAATACCTGTGATGGAATTTGATGTTGCCAAGGTTTTGGAGGAGCTTGAACCTACC 744
```

```
DB 12006 AAGAAGCTGCCCATCCAAGAGTTCCATCTGAGCGCGTCTCCAGSAGCTGGTCTGAAC 12065
QY 745 ATGACACAGTTTCATTTGATTTGTCATCTCTGTTGGATGTGACTATTTGATAGCATCAA 804
DB 12066 CAGGAGCAGTTTGTGGATCTGTGATCTCTGCTGGGTAGGACTACTGGGAGCATCCGT 12125
QY 805 GGTATCGGGGGCAACAGCAGCTCTGAAACTTATTCGTCAACATGGGTCCATAGAAGCATC 864
DB 12126 GGCATTTGGGCCCAAGCGGCTGTGGATCTCATCCAGAAACATAAGAGCATCAGAGATC 12185
QY 865 TTGAGAGATCTTAATAAGACAGATATCAAAATTCCTGAGGACTGGCCTTACCAAGAAGCT 924
DB 12186 GTAGGCGGCTGGACCCAGCAAGTACCCCGTTCCAGAGAACTGGCTCCAAAGGAAGCC 12245
QY 925 CGAGCTTTTCAAGGAGCCTAATG---TCACATTTGGATATTTCTCTGAGCTAAATGGACT 981
DB 12246 CAGCAGCTCTTCTTGGAGCCAGAACTACTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 12305
QY 982 GCACCTGATGAGGAGGCTCTCATAAAGTTTCTCTGTTAAAGATAATGGTTTCAACCAAGAT 1041
DB 12306 GAGCCAAATGAAGAAGAGTTGGTCAAAATTTATGTGTGGTGAAGACAGTTTCTCTGAAG 12365
QY 1042 CGGCTGCAAGGCCATAGAGAGATCAAAATCTGCCAAGATAAATCTGCCAAGGAAGA 1101
DB 12366 CCAATTCGAGTGGGTCAAGCGCTGAGTAAGAGCCGCGGAGCCAGCCAGCCAGGAGC 12425
QY 1102 CTCGAGCTCTTTTCAAGCCAACTGCCACACATCATGACACCGCTAAACGGAGGAGACT 1161
DB 12426 CTCGATGATTTTCAAGGTGACAGGCTCACTCTCCTCAGC-----TAAGCGC 12473
QY 1162 TCGGATAAACAACAGCAGGCTGGGCAACAGAAACAAAGCTGGTGGAAAGAAAGAA 1221
DB 12474 AAGGAGCCAGAACCCAGGGGCTGTCTAAGAGAAAGCAAGACTGGGGGCGGGAAG 12533
QY 1222 T 1222
DB 12534 T 12534

RESULT 12
BC027295 Mus musculus, clone MGC:28169 IMAGE:3985840, mRNA, complete cds.
LOCUS BC027295
DEFINITION BC027295.1 GI:20071070
ACCESSION BC027295
VERSION BC027295.1
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
```



Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 36 Row: g Column: 3  
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES

Location/Qualifiers  
 1..2102  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="C57BL/6J"  
 /clone="MGC:28169 IMAGE:3985840"  
 /tissue\_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."  
 /clone\_lib="NCI-CGAP\_Mam5"  
 /lab\_host="DH10B"  
 /note="vector: pCMV-SPORT6"  
 140..1282  
 /codon\_start=1  
 /product="unknown (protein for MGC:28169)"  
 /protein\_id="AAH27295.1"  
 /db\_xref="GI:20071071"  
 /translation="MGHGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYOFLL  
 AVRGDVLQNEGETTSHLMGFYRIRMGVAVPVYFDCKPQLKSGELARSE  
 RRAEKLQQAQAGMEVEEYKFKLVKTVKOHNDCKHLKSLPIYLDPEAE  
 ASCLAKAGKVAAYATEDMDCTFGSPVLMRLHTASEAKLPIDFHLRVLQELGL  
 NQOFVDLILGSDYCESIRIGIPRAVDLIQKHRSIEIIVRRDPSKYPPVNLH  
 KEAQLFLEPELVDPSEVLEKWPNEELVKPMCEKOFSEIRINSYKRLSKSRQG  
 STQGRDLDFEVKTVGLSSAKRKEPKRPAKKRATGGAGKFRKK"

BASE COUNT 536 a 502 c 600 g 464 t

Query Match 20.8%; Score 303.8; DB 10; Length 2102;  
 Best Local Similarity 56.2%; Pred. No. 5.2e-65;  
 Matches 641; Conservative 0; Mismatches 482; Indels 18; Gaps 3;

QY 85 ATGGGCAATCAAGGTTTACGAACTGCTGGCGGACATGCGCCCAAGCGCATGAAGGAG 144  
 DB 140 ATGGGAATTCACGGCTTCCCAAACTAATGCTGATGTGGCCCGCCAGTGGCATCCGTG 199  
 QY 145 CAGAAGTTCAGAGCTACTTTCGGCCGCAAAATCGCGTCGACGCGAGCATGAGCATATAC 204  
 DB 200 AATGACATCAAGAGCTACTTTGGTGCRAAGATGGCCATCGATGCCCTCATGAGCATCTAC 259  
 QY 205 CAGTCTCTGATTTAGTGGAGGACAGCATGGAACCTCTCAAAATGAAGCTGGTGA 264  
 DB 260 CAGTCTCTGATTTAGTGGAGGACAGCATGGAACCTCTCAAAATGAAGCTGGTGA 316  
 QY 265 GTCACTAGTCTATTCGAAGGAATGTTCAACCGGCAATAAGATTACTGGAAGCGGGAATC 324  
 DB 317 ACCACGAGCCACCTGATGGCATGTTCTACCGTACCATCCGATGATGGAGATGGCGTC 376  
 QY 325 AAGCCAGTTTATGTTTGTATGGCAAGCTCTCTGATATGAAGAACAAGAGCTTGTCTAAA 384  
 DB 377 AAGCTGTGTACTCTTTTGTATGGCAAGCTCTCTGATATGAAGAACAAGAGCTTGTCTAAA 436  
 QY 385 AGATACTCAAAAAGATGATGATCAACCAAGATCTGACTGAGCGGATAGAGGTAGGAGAT 444  
 DB 437 CGCAGTGAGAGGCGCGCGGAGGCTGAGAGCAACTGCAGCAGGCTCAGGAGGCTGGGATG 496  
 QY 445 AAGATGCGATTCAAAAATTTAGCAGAGGAGGCTGTAAGGTCACAGGCAACACACAGCAA 504  
 DB 497 GAGGAGGAGTGGAGAGTTCACCAAGAGGCTGCTGGAAGGTCAACCAACACACATGAT 556  
 QY 505 GATTGTAAGCGGTATTAAGACTTATFPGGGGTTCTCTGTTAGAGGCACTTCTGAAGCA 564  
 DB 557 GAGTGCAAAACCTGTGTAGGCTCATGCGGCATCCCTTACCTTGTATGCAACCCAGGAGGCA 616  
 QY 565 GAAGCAAGATGTCAGCCCTTTTCATATAAAGATAGGTTGCTGCTGTGTTCTTCAAGAT 624  
 DB 617 GAGCCAGCTGTGCTGCCCTGGCAAGGCTGGCAAGCTATGCTGCGGCCACCGAGGAC 676  
 QY 625 ATGGACTCCCTTACTTTTGGGGCTCCACCGTTCTTCTGCTCATTTAATGGATCAAGTTCC 684

DB 677 ATGACTGCCTCACTTTTGGCAGCCCGCTGCTAAATGCCACACTTAACGCGAGTGAGGCC 736  
 QY 685 AAGAAATACCTCTGATGGAATTTGATGTTGCCAAGGTTTTCGAGGAGCTTGAACCTACCC 744  
 DB 737 AAGAACTGCCATCCCAAGAGTTCCATCTGAGCCGCGCTCTCAGAGAGCTGGGTCTGAAC 796  
 QY 745 ATGACCAGTTCATTTGATTTGTCATCCCTGTTGATGAGTGAAGTGAATTTGATGATCAAA 804  
 DB 797 CAGGAGCAGTTTGTGGATCTGTGCATCTCTGCTGGGTAGGACTACTTCCAGAGGATCCGT 856  
 QY 805 GGTATCGGGGGGCAACAGCTCTGAACCTTATTGCTCAACATGCGTCCATAGAAAGCATC 864  
 DB 857 GGCATTGGGCCCAAGCGGCTGTGGATCTCATCCAGAAACATAAGAGCATCAGGAGATC 916  
 QY 865 TTGGAGATCTTTAATAAGACAGATATCAATTCCTGAGGAGTGGCCTTACCAAGAGCT 924  
 DB 917 GTGAGCGGCTGGACCCAGCAAGTACCCCGTTTCCAGAGAACTGGCTCCACAAGGAGGCC 976  
 QY 925 CGAGCTGTTCAGAGGAGCCTAATG--TCACATTTGGATATTCCTGAGCTAAATGGACT 981  
 DB 977 CAGCAGCTCTTCTTGGAGCCAGAACTGAGCCAGAGCTCTGTGAGCTGAAGTGGAGC 1036  
 QY 982 GCACCTGATGAGGAGGCTCATAGTTTCTCTGTTAAAGATAATGTTTCAACGAAGAT 1041  
 DB 1037 GAGCAATGAAGAGAGTTGGTCAATTTATGTTGGTGAAGAGCAGTTTCTGAAGAG 1096  
 QY 1042 CGGTGCAAAAGGCCATPAGAGAAGATCAATCTGCCAAGAAATAATTCGTCGCAAGGAAGA 1101  
 DB 1097 CGAATTCGAGTGGGGTCAAGCGGCTGAGTAAGAGCGCCAGGCGAGCACCAGGAGGCC 1156  
 QY 1102 CTCGAGTCTTTTCAAGCCAACTGCCACATCAGCACCGCTAAACCGAAGGAGACT 1161  
 DB 1157 CTCGATGATTTCTCAAGGTGACAGGCTCAGCTCTCTCTCAGC-----TAAGCGC 1204  
 QY 1162 TCGATTAACCAAGCAGGAGCTGCGACACAGAAACAAAGCTGCTGGTGAAGAGAA 1221  
 DB 1205 AAGAGCGCAACCAAGGGGCTGCTAAGAGAAAGCAAGAGCTGGGGGCGGGGAG 1264  
 QY 1222 T 1222  
 DB 1265 T 1265

RESULT 13

HSRAD2  
 LOCUS Homo sapiens mRNA for flap endonuclease-1.  
 DEFINITION  
 ACCESSION X76771  
 VERSION X76771.1 GI:1905802  
 KEYWORDS fln1 gene; flap endonuclease-1.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1757)  
 Wats.F.  
 Direct Submission  
 Submitted (16-DEC-1993) F. Wats, University of Sussex, School of  
 Biological Sciences, Falmer, Brighton BN1 9QG, UK  
 JOURNAL  
 REFERENCE  
 2 (bases 1 to 1757)  
 Murray, J.M., Tavassoli, M., al-Harithy, R., Sheldrick, K.S.,  
 Lehmann, A.R., Carr, A.M. and Wats, F.Z.  
 Structural and functional conservation of the human homolog of the  
 Schistosaccharomyces pombe rad2 gene, which is required for  
 chromosome segregation and recovery from DNA damage  
 Mol. Cell. Biol. 14 (7), 4878-4888 (1994)  
 MEDLINE 94277093  
 PUBMED 8007985  
 FEATURES  
 Location/Qualifiers  
 1..1757  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

```
gene 368..1510
CDS   /gene="fen1"
      368..1510
      /gene="fen1"
      /codon_start=1
      /product="flap endonuclease-1"
      /protein_id="CAA54166.1"
      /db_xref="GI:3980293"
      /translation="MGIGQLAKLIADVAPSAIRENDIKSVGRKVAIDASMSIYOFELI
AVROGDVIONEEGETTSHLMGFYETIRMMENGIKPYVDFDKPKSGSEIAKRSE
RRAEAKQLOQAQAEQVEKFTKRLVKYTKQHNDECKHLLSLMGIPVLDAPSAAE
ASCAALVKAGKYAAATEDMDCLTFGSPVLMRLHTASEAKKLPIOEPLHSRIQELGL
NQRFVLDLILLGSDYCESIRIGIKRAVDLIQKHSTIEIRLRLDNKVPVPENLWH
KEAHOLFLEPVDLPDESVELKWSPENEELIKFMCGEKOFSEERIRSGYKRLSKSQG
STQGRLDLDFKVTGSLSSAKRKEPEPKSKTKKAKTGAAGKFKRGK"
BASE COUNT 418 a 443 c 542 g 354 t
ORIGIN

Query Match      20.7%; Score 302.8; DB 9; Length 1757;
Best local similarity 55.0%; Pred. No. 8.9e-65;
Matches 661; Conservative 0; Mismatches 532; Indels 9; Gaps 3;

QY 85 ATGGGCATCAAGGGTTTGACGAACACTGCTGGCGGCAATGCGCCCAAGCGGATGAAGGAG 144
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 368 ATGGGAATTCAGAGCTGATGGCCGCAAACTAATGCTGATGTGGCGCCCGCCAGTCCGCGGAG 427

QY 145 CAGAAGTTCGAGAGCTACTTCGCGCGCAAAATCGCGCTCGACGCCAGCATGACATATAC 204
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 428 AATGACATCAAGAGCTACTTTGGCCGTAGGTAGGTGGCCATTGATGCTCTATGAGCATTTAT 487

QY 205 CAGTTCCTGATTTAGTTGAAGCAGCAGGATGGAACCTCTCACAAATGAAGCTGGTGA 264
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 488 CAGTTCCTGATTTGCTGTGTCGCCAG--GGTGGGGATGTCTGCAGAAATGAGGAGGTGAG 544

QY 265 GTCATGCTAGTATTTGCAAGAAATGTTCAACCGGACAAATGAAGATTACTGGAAGCGGGAATC 324
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 545 ACCACGAGCCACCTGATGGCCATGTTCTACCGCACCATTCGCTGATGGAGACGGCATC 604

QY 325 AAGCCAGTTATGTTTGTATGACGACGCCCTCTGATATGAAGAAACACAGAGCTTGTCTAA 384
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 605 AAGCCCGTGTATGCTTTGTATGACGACGCCCTCTGATGACGCGGATGCTGAGGAGTGGCCAAA 664

QY 385 AGATACTCAAAAGAGATGATGCAACCAAGATCTGACTGAGCGCAGTAGAGGTAGGAGAT 444
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 665 CCAGTGTAGCGCGGGCTGTAGGCGAGAGAGCAGCTGCAGCGGCTCAGGCTGTGGGGCC 724

QY 445 AAAGATGCGATTGGAATTTGACAGAGGACTGTAAAGGTCAACAGGCACACACACAGAA 504
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 725 GAGCAGAGGTGGAATAATTTCACTAAGCGGTGGTGAAGGTCTATGCTGCGGCTACCGAGAC 784

QY 505 GATTGTAACGGCTATTAGACATTTATGGGGTTTCCTGTTGTAGAGGCCACCTTCTGAAGCA 564
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 785 GAGTGCAAAACATCTGCTGAGCCCTCATGGGCATCCCTTATCTTGTATGACACCGAGTGAGCA 844

QY 565 GAAGCAGAAATGTGAGCGCCCTTTCATATAAAGATAGGTGTCCTGCTTGTCTGACAAGAT 624
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 845 GAGCCAGCTGTGCTGCGCTGGTGAAGGCTGGCAGAAGTCTATGCTGCGGCTACCGAGAC 904

QY 625 ATGAGCTCCCTTACTTTTGGGGCTCCACGGTTCCTTCGTCATTTAATGGATCCCAAGTTCC 684
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 905 ATGAGCTGCTCAGCTTCGCGACGCTGTGCTAATGCGACACCTGACTGCCAGTGAAGCC 964

QY 685 AAGAAAATACCTGTGATGGAATTTGATGTGGCAAGGTTTTGGAGGAGCTTGAACCTCAC 744
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 965 AAAAAGCTGCCAATCCAGGAATTTCCACCTGAGCGGATTTCTGCAGAGCTGGGCTGAAC 1024

QY 745 ATGACCAAGTTCATGATTTGTCATCCCTGTGATGATGATGATGATGATGATGATGATGATG 804
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 1025 CAGGACAGATTTGGATCTGTGATCTCCCTAGGACAGTACTGCTGAGAGATCCGG 1084

QY 805 GGTATCGGGGGGCAACAGCTCTGAAACCTTATTCGTCACATGCGTTCATGAAAGCATC 864
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 1085 GGTATGCGGCCCAAGCGGGCTGTGGACCTCATCCAGAGCACAAGAGCATCGAGGAGATC 1144
```

```
QY 865 TTGGAGAAATCTTAATAAAGACAGATATCAAAATTCCTGAGGACTGGCTTACCAAGAAGCT 924
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1145 GTGCGGCGACTTGACCCCAACAAGTACCTGTGCGAGAAAATTTGGCTCCACAAGGAGCT 1204

QY 925 CGACGCTGTTTCAAGGAGGACCTAATG---TCACATTTGGATATTCCTGAGCTAAATGGACT 981
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1205 CACCAAGCTCTTTCTTTGGAACTCTGAGGTGCTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 1264

QY 982 GCACCTGATGAGGAGGCTCTCATAAGTTTCTCTGTTAAAAGATAATGTTTCAACGAAGAT 1041
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1265 GAGCCAAATGAAGAAGAGCTGATCAAGTTTCATGTGTGGTGAAGAAGCAGTTCTCTGAGGAG 1324

QY 1042 CGGGTGACAAAGGCCATAGAGAGATCAAAATTCGCAAGAATAAATTCGTCGCAAGAGA 1101
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1325 CGAATCCGAGTGGGGTCAAGAGCTGATGATGAGCGCCCAAGGACGACCCAGGCGCGC 1384

QY 1102 CTCGAGTCTCTTTTCAAGCCCAACTGCGCAC---CACATCAGCACCCGCTTAAACGGAAGAG 1158
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1385 CTGGATGATTTCTTCAAGGTGACGGGCTCACTCTCTTCACTTAAGCGCAAGGAGCCAGAA 1444

QY 1159 ACTTCGGATATAAACAAGCAAGGAGCTGCGAACAAGAAACAAAGGCTGGTGGAAAGAG 1218
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1445 CCAAGGATCCACTAAGAAGAGGCAAGACTGGGCGACGAGGGAAGTTTAAAGGGA 1504

QY 1219 AAATAATCTTGGATCTGTGTACTACTACGACTAGCAAGACGCGTGGCTGATCATC 1278
    ||||| || || || || || || || || || || || || || || || || || || || ||
Db 1505 AAATAATGTGTTTCCCATATACCTCTTCAACCCAGAAATTTGGCGCTTGTACCC 1564

QY 1279 TT 1280
    ||
Db 1565 TT 1566

RESULT 14
BC000323 2031 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, flap structure-specific endonuclease 1, clone
DEFINITION MGC:8478 IMAGE:2821792, mRNA, complete cds.
ACCESSION BC000323
VERSION BC000323.1 GI:12653112
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Strausberg, R.
Direct Submission
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: qcapps-f@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.B., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, S., Stantripop, S., Thomas, P.J.,
Tongson, E.B., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
```

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 1 Row: 9 Column: 13  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 1905802.

## FEATURES

## source

1. .2031  
 /organism="Homo sapiens"  
 /db\_xref="LocusID:2237"  
 /db\_xref="taxon:9606"  
 /clone="MGC:8478 IMAGE:2821792"  
 /tissue-type="lung, small cell carcinoma"  
 /clone\_lib="NIH MGC-7"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 139. .1281

## CDS

/codon\_start=1  
 /product="flap structure-specific endonuclease 1"  
 /protein\_id="AAH00323.1"  
 /db\_xref="GI:12653113"  
 /translation="MGIQLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLI  
 AVROGDVLONEGETTSHLMGFYRTIRRMENIKPVYFDGKPPOLKSGELAKRSE  
 RRAEAKLOAOAGABOEVEKFTKRLVTKOHNDECKHLLSLQEPHLYLIDAPSBAE  
 ASCAALVAKGVYAAATEDMDCLFTFGSPVLMLHTASEAKKLPLOEPHLSRIQLBEL  
 NOEPVDLCILLSDYCESIRGIGPKRAVDLIQKHSIEIRVRLDRPKYVRENHLH  
 XEAHOLFEPVLDPESELKWSPEELIKFMCEKQFSEERIRSGVRLSKSRG  
 STQGRIDDFVKTSLSSAKRKEPEPKGTSKKAKTGAAGKFRKG"  
 533 a 465 c 589 g 444 t

## BASE COUNT

## ORIGIN

Query Match 20.7%; Score 302.8; DB 9; Length 2031;  
 Best Local Similarity 55.0%; Pred. No. 9, 2e-65;  
 Matches 661; Conservative 0; Mismatches 532; Indels 9; Gaps 3;  
 QY 85 ATGGGATCAAGGGTTGACGAACTCTGGCGGACATGCGCCCAAGCGCATGAGGAG 144  
 DB 139 ATGGGAATTCAGGCTGGCGCAAACTAAATGCTGATGTGGCCCCCAGCTGCCATCGCGGAG 198  
 QY 145 CAGAAGTCTGAGAGCTACTTCGGCGGCAAAATCGCGGTGCGAGCCGAGCATGAGCATATAC 204  
 DB 199 AATGACATCAAGAGCTACTTTGGCCGTAAGGTGGCCATTGATGCTCTATGACGATTTAT 258  
 QY 205 CAGTTCCTGATCTAGTGTGAGGAGGACGATGCAAACTCTCACAAATGAAGCTGGTAA 264  
 DB 259 CAGTTCCTGATCTGTTGCCAG---GGTGGGATGTGCTGCAGANTGAGGAGGAGTGA 315  
 QY 265 GTCACCTAGTCAATTTGCAAGGAATGTTCAACCGGACAAATAGATTACTTGAAGCGGGAATC 324  
 DB 316 ACCACGAGCCAGCTGATGGCGATGTTCTACCGCACCATTCGATGATGAGAGCAAGCGCATC 375  
 QY 325 AAGCCAGTCTATGTTTGTATGGCAAGCCTCTGATGATGAAGAAACAAGAGCTTCTGATAA 384  
 DB 376 AAGCCGCTGATGCTTTGATGCAAGCCGCCACAGCTCAAGTCAGGCGAGCTGCCCAA 435  
 QY 385 AGATCTCAAAAAGATGATGCAACCAAGATCTGACTGAGGCGAGTAGAGTAGAGAT 444  
 DB 436 CGCAGTGAGCGCGGCTGAGCGAGAGAGAGAGCTGCAGCAGGCTCAGGCTCTGGGCC 495  
 QY 445 AAGATCGGATTTGAAAATTTGACAGAGAGAGCTGTAAAGGTCACAAAGGCAACACACGAA 504  
 DB 496 GAGCAGAGGTGAAAATTTCACTAAGCGGCTGTGAAGGTCACCTAAGCAGCACAATGAT 555  
 QY 505 GATTGTAACGGCTATTAGACTTATGGGTTCTCTGTTGTAGAGGACCTTCTGAAGCA 564  
 DB 556 GAGTGCAAACTATGCTGAGCCCTCATGGGATCCCTTATCTGATGACCCAGTAGGAGCA 615  
 QY 565 GAAGCAGAATGTGACGCCCTTTGCATAAAGATGATGAGGTGTTCGCTGTTGCTTTCAGAAGAT 624  
 DB 616 GAGCCAGCTGTGTCGCTGTGGTGAAGGCTGGCAAAAGTCTATGCTCGGCTACCGAGGAG 675  
 QY 625 ATGACCTCCCTTACTTACTGGGCTCCACGCTTCCTTCGTCATTTAATGATCCAAGTCC 684  
 DB 676 ATGACTGCTCCTACCTTTCGCGAGCCCTGTGCTAATCGCACCTGACTGCCAGTGAAGCC 735

QY 685 AAGAAAATACCTGTGTGATGGAATTTGATGTGTCGAAGGTTTGGAGGAGCTTGAACCTACC 744  
 DB 736 AAAAACTGCCAATCCAGGAATTCACCTGAGCCGAGATTCTCGAGGAGCTGGGCTGAAC 795  
 QY 745 ATGAGCAGCTTCAATGATTTGTGATCCCTGTGTCAGTGTGACTATTGTGATGATGATCAAA 804  
 DB 796 CAGGAACAGTTTGGATCTGTGATCTCTGATCTCTAGGACAGTACTCTGTGAGAGTATCCGG 855  
 QY 805 GGTATCGGGGGGCAACAGCTCTGAACATTTATTCGTCAACATGGGTGCCATGAAGAAGCATC 864  
 DB 856 GGTATTGGGCCCAAGCGGCTGTGGACCTCATCCAGAGCAAGAGCATCGAGGAGATC 915  
 QY 865 TTGAGAGATCTTAAATAAGACAGATCAAAATTCCTGAGGACTGGCCTTACCAAGAAAGCT 924  
 DB 916 GTGCGGAGCTTGACCCCAACAGTACCTGTGCCAGAAATTTGGCTCCACAAGAGGCT 975  
 QY 925 CGACGCTGTTCAGGAGGCTTAATG---TCACATGGGATATTCCTGAGCTAAATGGACT 981  
 DB 976 CACCACTCTTCTTGAACCTGAGTGTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 1035  
 QY 982 GCACCTGATGAGGAGGCTCTCATAGTTTCTCTGTTAAAGATAAATGGTTTCAACGAAGAT 1041  
 DB 1036 GAGCCAAATGAAGAGAGCTGATCAAGTTTCAATGTGTGTTGAAAGCAGTTCTCTGAGGAG 1095  
 QY 1042 CGSGTGACAAAGCCATAGAGAGATCAAACTGTCACCAAGATTAATTCGTCGCAAGGAGA 1101  
 DB 1096 CGAATCGCAGTGGGTCAAGAGGCTGAGTAGAGCGCCAGGAGCAGCCAGGCGCGC 1155  
 QY 1102 CTCGAGTCTCTTTTCAAGCCAACTGCCAC---CACATAGCACCCGCTAAACGGGAAGAG 1158  
 DB 1156 CTGGATGATTTCTTCAAGGTGACCGGCTCACTCTCTTCAAGTCAAGGAGGAGCAGAA 1215  
 QY 1159 ACTTCGATATAAACAAGAGGAGCTGCGAACCAAGAAAGGCTGGTGGAAAGAG 1218  
 DB 1216 CCAAGGATCCACTAAGAGAGGCAAGAGCTGGGCGAGCGGAGGTTTAAAGGGGA 1275  
 QY 1219 AATAATCTTGGATGCTGTGATGATCAACATGACGACTAGAAAGAGCGGTGGCGTATAC 1278  
 DB 1276 AATAATGTTGTTTCCCATATATACCTCTTCAACCCAGAAATTTGCGGTCTTATACC 1335  
 QY 1279 TT 1280  
 DB 1336 TT 1337

RESULT 15  
 AF523117  
 LOCUS Homo sapiens flap structure-specific endonuclease 1 (FEN1) gene, complete cds.  
 DEFINITION AF523117  
 ACCESSION AF523117  
 VERSION AF523117.1 GI:21668122  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 6536)  
 AUTHORS Rieder, M.J., Livingston, R.J., Braun, A.C., Montoya, M.A., Chung, M.-W., Miyamoto, K.E., Nguyen, C.P., Nguyen, D.A., Poel, C.L., Robertson, P.D., Schackwitz, W.S., Sherwood, J.K., Witrak, L.A. and Nickerson, D.A.  
 DIRECT SUBMISSION  
 TITLE Submitted (20-JUN-2002) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA  
 JOURNAL To cite this work please use: NIHES-SNPs, Environmental Genome Project, NIHES ES15478, Department of Genome Sciences, Seattle, WA (URL: <http://eep.gs.washington.edu>).  
 COMMENT  
 FEATURES Location/Qualifiers  
 source 1. .6536  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 variation 103

```
/frequency="0.01"  
/replace="c"  
360  
variation  
/frequency="0.01"  
/replace="t"  
393.456  
repeat_region  
/rpt_family="L2"  
/rpt_type-dispersed  
504  
variation  
/frequency="0.01"  
/replace="t"  
974  
variation  
/frequency="0.01"  
/replace="t"  
982  
variation  
/frequency="0.01"  
/replace="c"  
995  
variation  
/frequency="0.40"  
/replace="a"  
1064.5624  
gene  
/gene="FEN1"  
Join(1064.1414,3727.5624)  
mRNA  
/gene="FEN1"  
/product="flap structure-specific endonuclease 1"  
1175  
variation  
/gene="FEN1"  
/frequency="0.07"  
/replace="t"  
1478.1628  
repeat_region  
/rpt_family="MIR"  
/rpt_type-dispersed  
1826.1971  
repeat_region  
/rpt_family="MIR"  
/rpt_type-dispersed  
2202  
variation  
/gene="FEN1"  
/frequency="0.02"  
/replace="g"  
2345.2510  
repeat_region  
/rpt_family="Alu"  
/rpt_type-dispersed  
2568.2787  
repeat_region  
/rpt_family="Alu"  
/rpt_type-dispersed  
2632  
variation  
/gene="FEN1"  
/frequency="0.01"  
/replace="a"  
2907.3006  
repeat_region  
/rpt_family="MER1_type"  
/rpt_type-dispersed  
3138  
variation  
/gene="FEN1"  
/frequency="0.01"  
/replace="g"  
3276.3319  
repeat_region  
/rpt_family="ERV1"  
/rpt_type-dispersed  
3690  
variation  
/gene="FEN1"  
/frequency="0.01"  
/replace="g"  
3748.4890  
CDS  
/gene="FEN1"  
/codon_start=1  
/product="flap structure-specific endonuclease 1"  
/protein_id="AM74238.1"  
/db_xref="GI:2168423"  
/translation="MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIQFLI  
AVRGQVLNDEGETSHLMGMFYETIRMENGIKPVYFDGPKPOLKSGELAKRSE  
RRAPAEQLQAQAQAEQVEKFKRLVKVTKQHNDECKHLILSLMGPYLDAPSEAE  
ASCAALVKAGKVAATATEDMDCLTFGSPVLMRHLTASEAKKLP1QEFHLSRLQELGL
```

```
NQQRVDCILGSDYCSIRIGIQPKRAVDLIQKHSTIEIVRLDLPNKYPVPENWLH  
KEAHLQLEPEVLDPSEVELKWESENEELIKFMCGEKQFSEERIRSGVKRLSKSRQG  
STQGRLLDDFFKVTGSLSSAKRKEPKGKTKKAKTGAAGFKRKG"  
4246.4405  
misc_feature  
/gene="FEN1"  
/note="region not scanned for variation"  
4920  
variation  
/gene="FEN1"  
/frequency="0.01"  
/replace="g"  
4973  
variation  
/gene="FEN1"  
/frequency="0.01"  
/replace="t"  
5213  
variation  
/gene="FEN1"  
/frequency="0.41"  
/replace="t"  
6108  
variation  
/frequency="0.03"  
/replace="t"  
6346  
variation  
/frequency="0.01"  
/replace="c"  
BASE COUNT 1701 a 1441 c 1706 g 1688 t  
ORIGIN  
Query Match 20.7%; Score 302.8; DB 9; Length 6536;  
Best local Similarity 55.0%; Pred. No. 1.1e-64;  
Matches 661; Conservative 0; Mismatches 532; Indels 9; Gaps 3;  
...Qy 85 ATGGGATCAAGGGTTTGACGAACCTGCTGGCGGACAAATGCGCCCAAGGCGCATGAAGAG 144  
Db 3748 ATGGGAATTCAGGCGCTGGCCAACTAATTGCTGATGTGGCCCCCGCATGCCATCCGGGAG 3807  
Qy 145 CAGAAAGTTCGAGAGCTACTTCGGCCGCAAAATCGCGCTGCGAGCCAGCATCATATAC 204  
Db 3808 AATGACATCAAGAGCTACTTTGGCCGTAAGGTGGCCATTTGATGCTTATGAGCATTTAT 3867  
Qy 205 CAGTCTCTGATTGTTGGAAGGACAGCATGGAACCTCTCAAAATGAAGCTGGTGAA 264  
Db 3868 CAGTCTCTGATTGTTGCGCCAG--GGTGGGATGTGCTGCAGAAATGAGAGGGTGAG 3924  
Qy 265 GTCACTAGTCAATTTGCAAGGAATGTTCAACGGGACAAATAAGATTACTGGAAGCGGAATC 324  
Db 3925 ACCACAGCCACCTGATGGCGCATGTTACCGCACCATTCGCATGATGGAGACGGCATC 3984  
Qy 325 AAGCCAGTTTATGTTTATGGCAAGCCTCTGATATGAGAAACAGAGCTTGCTAAA 384  
Db 3985 AAGCCCGTGTATGTTTATGGCAAGCCTCTGATGAGCCGACAGCTCAAGTCAGGCGAGCTGCCAAA 4044  
Qy 385 AGATACTCAAAAAGAGATGATCAACCAAGATCTGACTGAGGAGTAGAGTAGAGAT 444  
Db 4045 CGCAGTGAGCGGGGCTGAGGACAGAGAGCTGACGAGAGCTCAGGCTGCTGGGGCC 4104  
Qy 445 AAGATCCGATTCAAAATTCAGCAAGAGGAGCTGTAAAGTTCACAAGGACACACAA 504  
Db 4105 GAGCAGAGGTGGAATAATTCATAGCGGGTGTGAAGGTCACCTAAGCAGACCAATGAT 4164  
Qy 505 GATTGTAAAGGCTATTAAAGACTTATGGGGTTCCTGTTGTAGAGGACCTTCTGAAGCA 564  
Db 4165 GAGTGAACACATCTGCTGAGCCTCATGGCATCCCTTATCTTGTATGACCCAGTGAGGCA 4224  
Qy 565 GAAGCAGAATGTGACGCCCTTTGCATAAACGATAAGGTGTTGCTGCTCAGAGAT 624  
Db 4225 GAGCCAGCTGTGCTGCCCTGTTGAGGCTGGCAAGTCTATGCTGGGCTACCGAGGAC 4284  
Qy 625 ATGACTCCCTTACTTTTGGGCTCCACGGTTCCTTCCTCATTTAATGGATGCCAAGTCC 684  
Db 4285 ATGACTGCTCACCTTCGCGAGCCCTGTGTAATCGACACCTGACTGCCAGTGAAGCC 4344  
Qy 685 AAGAAAATACCTGTGATGAATTTTGTATGTTCGAAGGTTTGGAGAGAGCTTGAACCTCAC 744
```

```

Db 4345 AAAAAAGCTGCCAATCAGGAATCCACCTGAGCCGGATTCTGCAGGAGCTGGGCCTGAAC 4404
QY 745 ATGGACAGTTTCATTGATTTGTGCATCCCTGTGTGGATGTGACTATTGTGATAGCATCAAA 804
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4405 CAGGAACAGTTTGTGATCTGTGCATCCCTGTAGGACGTGACTACTGTGAGAGTATCCGG 4464
QY 805 GGTATCGGGGGGCAACACAGCTCTGAACCTTATTCGTCAACATGGGTCCATAGAAAAGCATC 864
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4465 GGTATTGGGGCCCAAGCGGCTGTGGACCTCATCCAGAGCAACAGAGCATCGAGGAGATC 4524
QY 865 TTGGAGAATCTTAAATAAGACAGATATCAAAATTCCTGAGACTGGCCTTACCAGAAGCT 924
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4525 GTCGGGGACTTGACCCCAACAAGTACCCTGTGCCAGAAAATTGGCTCCACAAGAGGCT 4584
QY 925 CGACGCTTGTTCAGGAGCCTAATG---TCACATTGGATATTCTTGAGCTAAATGGACT 981
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4585 CACCAAGCTCTTCTTGAACCTGAGGTGCTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 4644
QY 982 GCACCTGATGAGGAGGTCTCATAGTTTCCCTGTTAAAGATAATGGTTTCAACGAAGAT 1041
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4645 GAGCCAAATGAAGAAGAGCTGATCAAGTTTCATGTGTGTGAAAAGCAGTTCTCTGAGGAG 4704
QY 1042 CGSGTGACAAGGCCATAGAGAGATCAAAATCTGCCAAGAAATAATCTCGCAAGGAAGA 1101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4705 CGAATCCGCAAGTGGGTCAAGAGGCTGAGTAAGAGCCGCCAAGGCGAGCACCAGGCGCGC 4764
QY 1102 CTCGAGTCCCTTTTCAAGCAACTGCCAC---CACATCAGCACCGCTAAACCGGAAGGAG 1158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4765 CTGGATGATTTCTCAAGGTGACCGGCTCAGTCTCTTCAGCTAAGCGCAAGGAGCCAGAA 4824
QY 1159 ACTTCGGATAAACAGCAAGGAGCTGCGACAAGAAACAAAGGCTGGTGGAAAGAAG 1218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4825 CCCAAGGGATCCCACTAAGAGAAGGCAAGACTGGGGCGACGAGGGAAGTTTAAAGGGGA 4884
QY 1219 AATAATCTTTGGATGTTGATGTACAACTACGAAAGCAGCGGTGGGTGATCAG 1278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4885 AATAAATGTGTTCCCATTTACCTCTTACCCCAAGNATATTTGCCGCTCTTGATCCC 4944
QY 1279 TT 1280
Db 4945 TT 4946

```

Search completed: May 8, 2003, 21:02:13  
 Job time : 3813 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 19:38:47 ; Search time 360 Seconds  
(without alignments)  
9151.871 Million cell updates/sec

Title: US-09-805-311-1  
Perfect score: 1463  
Sequence: 1 cagcagaatagctcgccgccc.....aaaaaaaaaaaaaaaaaaaaa 1463

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues 4370478  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N_Geneseq_101002:*				
1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*			
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*			
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*			
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*			
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*			
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*			
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*			
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*			
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*			
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*			
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*			
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*			
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*			
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*			
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*			
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*			
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*			
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*			
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*			
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*			
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*			
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*			
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*			
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*			
SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	1463	100.0	1463-21	AAA27923 Maize Rad2/FEN-1 c
2	1421.8	97.2	1478-21	AAA27926 Maize Rad2/FEN-1 c
3	1409.8	96.4	1541-21	AAA27924 Maize Rad2/FEN-1 c
4	1350.2	92.3	1381-21	AAA27925 Maize Rad2/FEN-1 c
5	302.6	20.7	1144-20	AAAX02117 Human FEN-1 DNA fr
6	302.6	20.7	1144-20	AAAX02107 Human FEN-1 cDNA
7	276.6	18.9	2033-20	AAAX02111 Human FEN-1 genomi
8	276	18.9	1300-23	ABLI4287 Drosophila melanog
9	272.2	18.6	1930-20	AAAX02108 Mouse FEN-1 cDNA

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

10	270.2	18.5	1149	20	AAAX02109	Yeast FEN-1 cDNA
11	252.6	17.3	3362	23	ABLI4286	Drosophila melanog
12	173.2	11.8	5471	23	ABLI10094	Drosophila melanog
13	159.6	10.9	1023	18	AAAT76685	Pyrococcus furiosu
14	159.6	10.9	1023	19	AAV65840	Pyrococcus furiosu
15	159.6	10.9	1023	19	AAV53951	Nucleotide sequenc
16	152.8	10.4	1054	19	AAV53984	P. furiosus N-term
17	149.6	10.2	1115	19	AAV53992	P. furiosus N-term
18	139.6	9.5	510	22	AAK37479	Human bone marrow
19	139.6	9.5	510	22	AAI18286	Probe #8219 for ge
20	135.6	9.3	1032	20	AAAX31850	Pyrococcus heat re
21	134	9.2	349980	22	AAH41224	Pyrococcus abyssi
22	128	8.7	514	19	AAV53990	Nucleotide sequenc
23	126.4	8.6	1164	19	AAV53975	Nucleotide sequenc
24	125.6	8.6	1115	19	AAV53978	P. furiosus N-term
25	117.6	8.0	1729	23	AAAS83732	DNA encoding novel
26	110.2	7.5	889	19	AAV53974	Nucleotide sequenc
27	107.8	7.4	417	24	ABN77330	Human nucleasase-lik
28	98.6	6.7	1011	19	AAV65913	Archaeoglobus fulg
29	98.6	6.7	1011	19	AAV54009	Nucleotide sequenc
30	97.6	6.7	1164	19	AAV53979	M. jannaschii N-te
31	94.8	6.5	386	19	AAV53979	Human foetal liver
32	90	6.2	546	22	ABA63065	Human brain expres
33	90	6.2	546	22	AAK11490	Human bone marrow
34	90	6.2	546	22	AAK37269	Human bone marrow
35	90	6.2	546	22	AAI18098	Probe #8031 for ge
36	90	6.2	546	24	ABS11263	Human genome-deriv
37	89.2	6.1	981	18	AAV76682	Methanococcus jann
38	89.2	6.1	981	19	AAV53948	Nucleotide sequenc
39	89.2	6.1	1664976	19	AAV21209	Methanococcus jann
40	86.6	5.9	245	22	AAK50354	Human bone marrow
41	86.6	5.9	245	22	AAI27416	Probe #17349 for g
42	86	5.9	981	19	AAV65837	Methanococcus jann
43	76.2	5.2	777	19	AAV54010	Nucleotide sequenc
44	74.6	5.1	840	19	AAV53977	Nucleotide sequenc
45	69.4	4.7	987	19	AAV54011	Nucleotide sequenc

ALIGNMENTS

RESULT 1  
AAA27923  
ID AAA27923 standard; cDNA; 1463 BP.  
XX  
XX AAA27923;  
XX  
XX 12-SEP-2000 (first entry)  
DT  
XX Maize Rad2/FEN-1 cDNA.  
DE  
XX Maize; Rad2/FEN-1; transgenic plant; male sterile plant;  
KW endonuclease; exonuclease; DNA repair; gene targeting; ss.  
XX  
XX Zea mays.  
OS  
XX  
XX Key Location/Qualifiers  
FH 85..1224  
FT /\*tag= a  
XX  
XX WO200036109-A1.  
XX  
XX 22-JUN-2000.  
XX  
XX 16-NOV-1999; 99WO-US27147.  
XX  
XX 15-DEC-1998; 98US-0112332.  
XX  
XX (PION-) PIONEER HI-BRED INT INC.  
XX  
XX Mahajan PB;  
XX  
XX WPI; 2000-452026/39.

DR P-PSDB; AAY95307.

XX Maize RAD2/FEN-1 nucleic acids and proteins useful for modulating DNA  
PT recombination and repair in transgenic plants, e.g. for gene targeting  
PT and the production of male sterile plants -  
XX  
PS Claim 1; Page 69-71; 85pp; English.  
XX

CC The present sequence is that of maize cDNA coding for RAD2/FEN-1  
CC (see AAY95307). The corresponding RNA was isolated from immature  
CC ear tissue from 2 ears of a B73 maize line. Rad2/FEN-1 is a  
CC structure specific endonuclease which under certain conditions also  
CC acts as an exonuclease. Rad2/FEN-1 nucleic acids can be used to  
CC produce the Rad2/FEN-1 polypeptides in transgenic plant cells.  
CC The protein is involved in the regulation of DNA repair and  
CC recombination in plant systems and therefore may be used for  
CC improving gene targeting during further recombinant DNA protocols  
CC involving plants. RAD2/FEN-1 endonucleolytic activity is essential  
CC in DNA replication and nucleotide excision and repair reactions.  
CC The exolytic activity is involved in double strand break repair and  
CC end joining. The protein is also useful in strand exchange  
CC reactions during homologous recombination. These functions may be  
CC useful in gene targeting and in the production of male sterile  
CC plants. The efficacy of gene targeting can be improved by the  
CC overexpression of exogenous Rad2/FEN-1 while male sterile plants can  
CC be produced by the down regulation of Rad2/FEN-1 expression.

XX SQ Sequence 1463 BP; 466 A; 292 C; 361 G; 344 T; 0 other;

Query Match 100.0%; Score 1463; DB 21; Length 1463;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CACGAGAATAGCTCGGGCGCGGTTCTTGGCCACTCCGCTCAGCGCGCGCCGCCA	60
DB	1	CACGAGNATAGCTCGGGCGCGGTTCTTGGCCACTCCGCTCAGCGCGCGCCGCCA	60
QY	61	CCGCGCACAGCCCGCCGACGAGATGGCATCAAGGGTTTGACGAAACTGCTGGCGGAC	120
DB	61	CCGCGCACAGCCCGCCGACGAGATGGCATCAAGGGTTTGACGAAACTGCTGGCGGAC	120
QY	121	AATGCGCCCAAGCGGATGAAGGACGAGATGTCGAGAGCTACTTCGGCGCGCAAAATCGCC	180
DB	121	AATGCGCCCAAGCGGATGAAGGACGAGATGTCGAGAGCTACTTCGGCGCGCAAAATCGCC	180
QY	181	GTGACGCCAGCATGACATATACAGTTCCTGATTTGATGGAGACAGGATGAA	240
DB	181	GTGACGCCAGCATGACATATACAGTTCCTGATTTGATGGAGACAGGATGAA	240
QY	241	ACTCTCACAAATGAAGCTGGTGAAGTCACTAGTCAATTTGCAAGGAATGTTCAACCGGACA	300
DB	241	ACTCTCACAAATGAAGCTGGTGAAGTCACTAGTCAATTTGCAAGGAATGTTCAACCGGACA	300
QY	301	ATAAGATTACTGGAAGCGGAATCAAGCCAGTTTATGTTTTGATGGCAAGCCCTCCTGAT	360
DB	301	ATAAGATTACTGGAAGCGGAATCAAGCCAGTTTATGTTTTGATGGCAAGCCCTCCTGAT	360
QY	361	ATCAAGAAACAGAGCTTCTTAAAGATACCTCAAAAGAGATGATGCAACCAAGATCTG	420
DB	361	ATCAAGAAACAGAGCTTCTTAAAGATACCTCAAAAGAGATGATGCAACCAAGATCTG	420
QY	421	ACTGAGGCAGTAGAGTAGGATGAAGATGCGATTGAAAAATTGAGCAAGAGGACTGTA	480
DB	421	ACTGAGGCAGTAGAGTAGGATGAAGATGCGATTGAAAAATTGAGCAAGAGGACTGTA	480
QY	481	AAGGTACAGAGGCAACACAGGAATGTTAAACGGCTATTAAAGACTTATGGGGTTCCT	540
DB	481	AAGGTACAGAGGCAACACAGGAATGTTAAACGGCTATTAAAGACTTATGGGGTTCCT	540
QY	541	GTGTAGAGGCACCTTCTCAAGCAGAGCATGTCAGCCCTTTGATAAACGATAAG	600
DB	541	GTGTAGAGGCACCTTCTCAAGCAGAGCATGTCAGCCCTTTGATAAACGATAAG	600

QY	601	GTGTTGCGTGTGCTTCAGAGATATGAGCTCCCTTACTTTTGGGGCTCCACGGTTCCCTT	660
DB	601	GTGTTGCGTGTGCTTCAGAGATATGAGCTCCCTTACTTTTGGGGCTCCACGGTTCCCTT	660
QY	661	CGTCATTTAATGATCCCAAGTTTCCAGAAATACCTGTGATGGAATTTGATGTTGCCAAG	720
DB	661	CGTCATTTAATGATCCCAAGTTTCCAGAAATACCTGTGATGGAATTTGATGTTGCCAAG	720
QY	721	GTGTTGAGGAGCTTGAACCTCACCATGGACGAGTTCATTTGATTTGTGCATCCTGTGTGGA	780
DB	721	GTGTTGAGGAGCTTGAACCTCACCATGGACGAGTTCATTTGATTTGTGCATCCTGTGTGGA	780
QY	781	TGTGACTATTGTAGATCAAGATCAAGGTATCGGGGGCAACAGCTCTGAACATTTATTCGT	840
DB	781	TGTGACTATTGTAGATCAAGGTATCGGGGGCAACAGCTCTGAACATTTATTCGT	840
QY	841	CAACATGGTCCATAGAAGCATCTTGGAGATCTTTAATAAAGACAGATATCAAAATTCCT	900
DB	841	CAACATGGTCCATAGAAGCATCTTGGAGATCTTTAATAAAGACAGATATCAAAATTCCT	900
QY	901	GAGGACTGCGCTTCCCAAGAGCTCGACCTTGTTCAGGAGCCTAATGTCACATTCGAT	960
DB	901	GAGGACTGCGCTTCCCAAGAGCTCGACCTTGTTCAGGAGCCTAATGTCACATTCGAT	960
QY	961	ATTCTGTAGCTAAAATGGAGCTGCACCTGATGAGGAGGCTCTCATAAGTTTCTCGTAAAA	1020
DB	961	ATTCTGTAGCTAAAATGGAGCTGCACCTGATGAGGAGGCTCTCATAAGTTTCTCGTAAAA	1020
QY	1021	GATAATGTTTCAACGAGATCGGGTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAG	1080
DB	1021	GATAATGTTTCAACGAGATCGGGTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAG	1080
QY	1081	ATAAATCTGTCGAAGGAGCTCGAGTCTTTTCAAGCCCAACTGCCACCATCAGCA	1140
DB	1081	ATAAATCTGTCGAAGGAGCTCGAGTCTTTTCAAGCCCAACTGCCACCATCAGCA	1140
QY	1141	CGCTAAACGGAAGGAGCTTCGGATATAACAGCAGGAGGCTGGCAACAGAAAAACA	1200
DB	1141	CGCTAAACGGAAGGAGCTTCGGATATAACAGCAGGAGGCTGGCAACAGAAAAACA	1200
QY	1201	AAGGCTGGTGGAAAGAAATAATCTTTGGATGCTGATGTACAACTACGACTACGAAAG	1260
DB	1201	AAGGCTGGTGGAAAGAAATAATCTTTGGATGCTGATGTACAACTACGACTACGAAAG	1260
QY	1261	CAGCGTGGCGTGATCAGTTCGCTTAGATTTTAACTCCCTGTTTAACTCAGAGCTTT	1320
DB	1261	CAGCGTGGCGTGATCAGTTCGCTTAGATTTTAACTCCCTGTTTAACTCAGAGCTTT	1320
QY	1321	GGTAAAGTTTCTCATGTTTCAAGCTGGGTAAAGTTAGTTGTTTGAAGAGATTGGTG	1380
DB	1321	GGTAAAGTTTCTCATGTTTCAAGCTGGGTAAAGTTAGTTGTTTGAAGAGATTGGTG	1380
QY	1381	TACCAAGTAAACAACTTATCGCTGTTTTTACTCTTTTGTCTTTGAGTAAAGAAAAAAA	1440
DB	1381	TACCAAGTAAACAACTTATCGCTGTTTTTACTCTTTTGTCTTTGAGTAAAGAAAAAAA	1440
QY	1441	AAAAAATAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1463
DB	1441	AAAAAATAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1463

RESULT 2

AAA27926

ID AAA27926 standard; cDNA; 1478 BP.

XX AAA27926;

XX 12-SEP-2000 (first entry)

XX Maize Rad2/FEN-1 cDNA.

XX Maize; Rad2/FEN-1; transgenic plant; male sterile plant;  
KW endonuclease; exonuclease; DNA repair; gene targeting; ss.



XX OS      Zea mays.  
 XX FH      Key      Location/Qualifiers  
 XX FT      CDS      97..1236  
 XX FT           /\*tag= a  
 XX PN      WO200036109-A1.  
 XX PD      22-JUN-2000.  
 XX PF      16-NOV-1999; 99WO-US27147.  
 XX PR      15-DEC-1998; 98US-0112332.  
 XX PA      (PION-) PIONEER HI-BRED INT INC.  
 XX PI      Mahajan PB;  
 XX      WPI; 2000-452026/39.  
 DR      P-PSDB; AAY95310.  
 XX        
 PT      Maize RAD2/FEN-1 nucleic acids and proteins useful for modulating DNA  
 PT      recombination and repair in transgenic plants, e.g. for gene targeting  
 PT      and the production of male sterile plants  
 XX        
 PS      Example 1; Page 77-79; 85pp; English.  
 XX        
 CC      The present sequence is that of maize cDNA coding for RAD2/FEN-1  
 CC      (see AAY95310). The cDNA was isolated from a library prepared  
 CC      from B73 line seed vitreous endosperm RNA. Rad2/FEN-1 is a  
 CC      structure specific endonuclease which under certain conditions also  
 CC      acts as an exonuclease. Rad2/FEN-1 nucleic acids can be used to  
 CC      produce the Rad2/FEN-1 polypeptides in transgenic plant cells.  
 CC      The protein is involved in the regulation of DNA repair and  
 CC      recombination in plant systems and therefore may be used for  
 CC      improving gene targeting during further recombinant DNA protocols  
 CC      involving plants. RAD2/FEN-1 endonucleolytic activity is essential  
 CC      in DNA replication and nucleotide excision and repair reactions.  
 CC      The exolytic activity is involved in double strand break repair and  
 CC      end joining. The protein is also useful in strand exchange  
 CC      reactions during homologous recombination. These functions may be  
 CC      useful in gene targeting and in the production of male sterile  
 CC      plants. The efficacy of gene targeting can be improved by the  
 CC      overexpression of exogenous Rad2/FEN-1 while male sterile plants can  
 CC      be produced by the down regulation of Rad2/FEN-1 expression.  
 XX        
 SQ      Sequence 1478 BP; 463 A; 302 C; 365 G; 348 T; 0 other;  
  
 Query Match      97.2%; Score 1421.8; DB 21; Length 1478;  
 Best Local Similarity      98.5%; Pred. No. 0;  
 Matches 1435; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
  
 QY      7    AATAGCTCGCGGCGGTTCTTGGCCACTCCGGCTCAGCGCGCGCGCCGACCCGCC 66  
 DB      19    AATAGCTCGCGGTCGGGTTCTTGGCCACTCCGGCTCAGCGCGCGCCGACCCGCC 78  
  
 QY      67    ACAGCCGCCGACAGAGATGGGCATCAAGGGTTGACGAACTGCTGCGGACAAATGCG 126  
 DB      79    ACAGCCGCCGACAGAGATGGGCATCAAGGGTTGACGAACTGCTGCGGACAAATGCG 138  
  
 QY      127    CCCAAGCGGATGAAGAGCAGAGATTTCGAGAGCTACTTCGGCGGCAAAATCGCGCTCGAC 186  
 DB      139    CCCAAGCGGATGAAGAGCAGAGATTTCGAGAGCTACTTCGGCGGCAAAATCGCGCTCGAC 198  
  
 QY      187    GCCAGCATGAGCATATACCAAGTTCCTGATTTAGTTGGAGGACAGCGATGGAATCTCTC 246  
 DB      199    GCCAGCATGAGCATATACCAAGTTCCTGATTTAGTTGGAGGACAGCGATGGAATCTCTC 258  
  
 QY      247    ACAATCAAGCTGCTGAAGTCACTAGTCAATTCGAAGGATGTTCAACCGGCAATAAGA 306  
 DB      259    ACAATCAAGCTGCTGAAGTCACTAGTCAATTCGAAGGATGTTCAACCGGCAATAAGA 318

QY      307    TTACTGGAAGCGGAATCAAGCCAGTTTATGTTTTTGTATGSCAAGCCTCCTGATATGAG 366  
 DB      319    TTACTGGAAGCGGAATCAAGCCAGTTTATGTTTTTGTATGSCAAGCCTCCTGATATGAG 378  
  
 QY      367    AAACAAGAGCTTGTCTAAAAGATCTCAAAAGAGATGATGCAACCAAGATCTGACTGAG 426  
 DB      379    AAACAAGAGCTTGTCTAAAAGATCTCAAAAGAGATGATGCAACCAAGATCTGACTGAG 438  
  
 QY      427    GCAGTAGAGTAGGATGATGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATG 486  
 DB      439    GCAGTAGAGTAGGATGATGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATG 498  
  
 QY      487    ACAAGGCAACACAGAGATGTTAAAGCGCTATTAGACTTATGAGGCTTCTGTTGTA 546  
 DB      499    ACAAGGCAACACAGAGATGTTAAAGCGCTATTAGACTTATGAGGCTTCTGTTGTA 558  
  
 QY      547    GAGCACCCTTCTGAAGCAGAGAGATGTCAGCCCTTTGCATATAAAGATGATGATGTC 606  
 DB      559    GAGCACCCTTCTGAAGCAGAGAGATGTCAGCCCTTTGCATATAAAGATGATGATGTC 618  
  
 QY      607    GCTGTGCTTCAGAGATATGAGCTCCCTTACTTTTGGGGCTCCACGTTCTCTTCGTCAT 666  
 DB      619    GCTGTGCTTCAGAGATATGAGCTCCCTTACTTTTGGGGCTCCACGTTCTCTTCGTCAT 678  
  
 QY      667    TTAATGGATCAAGTTCACAGAAATACCTGCTGATGGAATTTGATTTGCCAAGGTTTG 726  
 DB      679    TTAATGGATCAAGTTCACAGAAATACCTGCTGATGGAATTTGATTTGCCAAGGTTTG 738  
  
 QY      727    GAGGAGCTTGAACCTCACCATGGACAGTTTCATTTGTTGTCATCCTGTTGGATGTCAC 786  
 DB      739    GAGGAGCTTGAACCTCACCATGGACAGTTTCATTTGTTGTCATCCTGTTGGATGTCAC 798  
  
 QY      787    TATTGTGATAGCATCAAGGATTCGGGGGCAACAGCTCTGAACTTATTTCGTCACAT 846  
 DB      799    TATTGTGATAGCATCAAGGATTCGGGGGCAACAGCTCTGAACTTATTTCGTCACAT 858  
  
 QY      847    GGGTCCATAGAAAGCATCTTGGAGAATCTTAATAAGACAGATATCAAAATTCCTTGAGGAC 906  
 DB      859    GGGTCCATAGAAAGCATCTTGGAGAATCTTAATAAGACAGATATCAAAATTCCTTGAGGAC 918  
  
 QY      907    TGGCTTACCAGAGCTCGACGCTTGTTCAGGAGCTTAATGTCACATTCGATATTCCT 966  
 DB      919    TGGCTTACCAGAGCTCGACGCTTGTTCAGGAGCTTAATGTCACATTCGATATTCCT 978  
  
 QY      967    GAGCTAAATGGACTGACCTGATGAGGAGGCTCTATAAGTTTCTGTTGTTAAAGATAAT 1026  
 DB      979    GAGCTAAATGGACTGACCTGATGAGGAGGCTCTATAAGTTTCTGTTGTTAAAGATAAT 1038  
  
 QY      1027    GGTTCACAGAGATCGGGTGCACAAAGGCCATAGAGAGATCAAAATCTGCCAAGATAAA 1086  
 DB      1039    GGTTCACAGAGATCGGGTGCACAAAGGCCATAGAGAGATCAAAATCTGCCAAGATAAA 1098  
  
 QY      1087    TCGTCGCAAGGAAGACTCGAGTCCCTTTTCAAGCCAACTGCCACCATCATCAGCACCGCTA 1146  
 DB      1099    TCGTCGCAAGGAAGACTCGAGTCCCTTTTCAAGCCAACTGCCACCATCATCAGCACCGCTA 1158  
  
 QY      1147    AAACGGAAGGAGACTTCGGATAAACAAGCAGCGCTGCGAACAAGAAACAAAGGCT 1206  
 DB      1159    AAACGGAAGGAGACTTCGGATAAACAAGCAGCGCTGCGAACAAGAAACAAAGGCT 1218  
  
 QY      1207    GGTGAAAGAAAGAAATAATCTTGGATGCTTGTATGATGATGATGATGATGATGATGATG 1266  
 DB      1219    GGTGAAAGAAAGAAATAATCTTGGATGCTTGTATGATGATGATGATGATGATGATGATG 1278  
  
 QY      1267    TGGCGTGATGACTTCGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1326  
 DB      1279    TGGCGTGATGACTTCGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1338  
  
 QY      1327    AGTTTGTCTCATGTTTCAAGCTGGGGTAAAGTTAGTTGTTTCAAGAGATTCGTTGATACAA 1386  
 DB      1339    AGTTTGTCTCATGTTTCAAGCTGGGGTAAAGTTAGTTGTTTCAAGAGATTCGTTGATACAA 1398  
  
 QY      1387    GTAACAAACATTAATGCGTGTGTTTTTACTTCTTGTGAGTAAAAAAGAAAAA 1446

|||||  
Db 1399 GTAACAAACTTATGCTGTTTTTACTCTTGTCTTGAAGTATGATGCCAGTAAA 1458  
| | | | |

QY 1447 AAAAAAAAAAAAAA 1463  
| | | | |

Db 1459 AAAAAAAAAAAAAA 1475  
| | | | |

## RESULT 3

AAA27924

ID AAA27924 standard; cDNA; 1541 BP.

XX AC AAA27924;

XX DE 12-SEP-2000 (first entry)

XX DE Maize Rad2/FEN-1 cDNA.

XX Maize; Rad2/FEN-1; transgenic plant; male sterile plant;  
KW endonuclease; exonuclease; DNA repair; gene targeting; ss.  
XX Zea mays.

XX Key Location/Qualifiers  
FH 79..1218  
CDS /\*tag= a

XX WO200036109-A1.

XX PD 22-JUN-2000.

XX PF 16-NOV-1999; 99WO-US27147.

XX PR 15-DEC-1998; 98US-0112332.

XX PA (PION-) PIONEER HI-BRED INT INC.

XX PI Mahajan PB;

XX WPI; 2000-452026/39.

XX DR P-PSDB; AAY95308.

PT Maize RAD2/FEN-1 nucleic acids and proteins useful for modulating DNA  
PT recombination and repair in transgenic plants, e.g. for gene targeting  
PT and the production of male sterile plants -  
XX Example 1; Page 73-74; 85pp; English.

XX The present sequence is that of maize cDNA (ATCC PTA-533) coding  
CC for RAD2/FEN-1 (see AAY95308). The corresponding RNA was isolated  
CC from a B73 line seedling after a 10 day drought, heat shocked for  
CC 10 hr, and allowed to recover under normal conditions. Rad2/FEN-1  
CC is a structure specific endonuclease which under certain conditions  
CC also acts as an exonuclease. Rad2/FEN-1 nucleic acids can be used  
CC to produce the rad2/FEN-1 polypeptides in transgenic plant cells.  
CC The protein is involved in the regulation of DNA repair and  
CC recombination in plant systems and therefore may be used for  
CC improving gene targeting during further recombinant DNA protocols  
CC involving plants. RAD2/FEN-1 endonucleolytic activity is essential  
CC in DNA replication and nucleotide excision and repair reactions.  
CC The exolytic activity is involved in double strand break repair and  
CC end joining. The protein is also useful in strand exchange  
CC reactions during homologous recombination. These functions may be  
CC useful in gene targeting and in the production of male sterile  
CC plants. The efficacy of gene targeting can be improved by the  
CC overexpression of exogenous Rad2/FEN-1 while male sterile plants can  
CC be produced by the down regulation of Rad2/FEN-1 expression.

XX Sequence 1541 BP; 473 A; 308 C; 377 G; 383 T; 0 other;

## Query Match

Best Local Similarity 96.4%; Score 1409.8; DB 21; Length 1541;

Matches 1411; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 19 CGCGGTTTCTTGGCCACTCCGCTCAGCGCGCCGCCACCGCCACAGCGCGCA 78  
| | | | |  
Db 13 CGCGGTTTCTTGGCCACTCCGCTCAGCGCGCCGCCACCGCCACAGCGCGCA 72  
| | | | |  
QY 79 GACGAGATGGGCATCAAGGGTTTGACGAACTGCTGGCGGACAATGGCCCAAGGCGATG 138  
| | | | |  
Db 73 GACGAGATGGGCATCAAGGGTTTGACGAACTGCTGGCGGACAATGGCCCAAGGCGATG 132  
| | | | |  
QY 139 AAGGAGCAGAAGTTTCGAGAGCTACTTCGGCGGCAAAATCCCGTCGACGCCACATGAGC 198  
| | | | |  
Db 133 AAGGAGCAGAAGTTTCGAGAGCTACTTCGGCGGCAAAATCCCGTCGACGCCACATGAGC 192  
| | | | |  
QY 199 ATATACCAAGTTCTCTGATTGTGTTGGAAGGACAGGCACTGAAAATCTCTCAAAATGAAGCT 258  
| | | | |  
Db 193 ATATACCAAGTTCTCTGATTGTGTTGGAAGGACAGGCACTGAAAATCTCTCAAAATGAAGCT 252  
| | | | |  
QY 259 GGTGAAGTCACTAGTCAATTTGCAAGGAATGTTCAACCGGACAATAAGATTACTGGAGCG 318  
| | | | |  
Db 253 GGTGAAGTCACTAGTCAATTTGCAAGGAATGTTCAACCGGACAATAAGATTACTGGAGCG 312  
| | | | |  
QY 319 GGAATCAAGCCAGTTTATGTTTGTGATGGCAAGCCTCCTGATATGAAGAAACAAGAGCTT 378  
| | | | |  
Db 313 GGAATCAAGCCAGTTTATGTTTGTGATGGCAAGCCTCCTGATATGAAGAAACAAGAGCTT 372  
| | | | |  
QY 379 GCTAAAAGATATCAAAAAGAGATGATCAACCAAGATCTGACTGAGGCGAGTAJAGTA 438  
| | | | |  
Db 373 GCTAAAAGATATCAAAAAGAGATGATCAACCAAGATCTGACTGAGGCGAGTAJAGTA 432  
| | | | |  
QY 439 GGAGATAAAGATGCGATTGAAAAATGAGCAAGAGGACTCTAAAGGTCAACAAGCAACAC 498  
| | | | |  
Db 433 GGAGATAAAGATGCGATTGAAAAATGAGCAAGAGGACTCTAAAGGTCAACAAGCAACAC 492  
| | | | |  
QY 499 AACGAAGATTGTAACGCGCTATTAAAGACTTATGGGGTTCCTGTTGAGAGGCACTTCT 558  
| | | | |  
Db 493 AACGAAGATTGTAACGCGCTATTAAAGACTTATGGGGTTCCTGTTGAGAGGCACTTCT 552  
| | | | |  
QY 559 GAAGCAGAAGCAGAATGTGAGCCCTTTGCTATAACGATAAAGGTGTTGCTGTGTTCA 618  
| | | | |  
Db 553 GAAGCAGAAGCAGAATGTGAGCCCTTTGCTATAACGATAAAGGTGTTGCTGTGTTCA 612  
| | | | |  
QY 619 GAAGATATGGATCCCTTACTTTTGGGGCTCCAGGTTCTCTGCTCATTTAATGGATCA 678  
| | | | |  
Db 613 GAAGATATGGATCCCTTACTTTTGGGGCTCCAGGTTCTCTGCTCATTTAATGGATCA 672  
| | | | |  
QY 679 AGTTCCAAAGAAATACCTGTGATGGAATTTGATGTTGCCAAGGTTTTGGAGGAGCTTGA 738  
| | | | |  
Db 673 AGTTCCAAAGAAATACCTGTGATGGAATTTGATGTTGCCAAGGTTTTGGAGGAGCTTGA 732  
| | | | |  
QY 739 CTCACCATGGACAGTTTCATTTGTTGTCATCTGTGTCATGTCATTTGTCATGTCATG 798  
| | | | |  
Db 733 CTCACCATGGACAGTTTCATTTGTTGTCATCTGTGTCATGTCATTTGTCATGTCATG 792  
| | | | |  
QY 799 ATCAAAGTATCGGGGGGCAACACAGCTCTGAAACTTATTGTCGAACATGGGTCATAGAA 858  
| | | | |  
Db 793 ATCAAAGTATCGGGGGGCAACACAGCTCTGAAACTTATTGTCGAACATGGGTCATAGAA 852  
| | | | |  
QY 859 AGCATCTGGAGATCTTAATAAGACAGATATAAATCTCTGAGACTGGCTTACCAA 918  
| | | | |  
Db 853 AGCATCTGGAGATCTTAATAAGACAGATATAAATCTCTGAGACTGGCTTACCAA 912  
| | | | |  
QY 919 GAAGCTCGACGCTTGTTCAGAGGAGCCCTAATGTGCATATTCCTGAGCTAAATGG 978  
| | | | |  
Db 913 GAAGCTCGACGCTTGTTCAGAGGAGCCCTAATGTGCATATTCCTGAGCTAAATGG 972  
| | | | |  
QY 979 ACTGCACCTGTAGAGGAGGCTCTCATAGTTTCTCGTAAAGATAATGTTTCAACGAA 1038  
| | | | |  
Db 973 ACTGCACCTGTAGAGGAGGCTCTCATAGTTTCTCGTAAAGATAATGTTTCAACGAA 1032  
| | | | |  
QY 1039 GATCGGTTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAAATAATCTGTCGAAGA 1098  
| | | | |  
Db 1033 GATCGGTTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAAATAATCTGTCGAAGA 1092  
| | | | |



QY 889 TATCAATTCCTGAGGACTGGCCCTTACCAAGAGCTCGACGCTTGTTCAGAGGCGCTAAT 948  
 Db 841 TATCAATTCCTGAGGACTGGCCCTTACCAAGAGCTCGACGCTTGTTCAGAGGCGCTAAT 900  
 QY 949 GTCAATTCGGATATTCCTGAGCTAAATGGACTGACCTGATGAGGAGGCTCTATAAGT 1008  
 Db 901 GTCAATTCGGATATTCCTGAGCTAAATGGACTGACCTGATGAGGAGGCTCTATAAGT 960  
 QY 1009 TTCCTGGTAAAGATAAATGGTTTCAACGAAGATCGGGTGACAAAGCCATAGAGAAGATC 1068  
 Db 961 TTCCTGGTAAAGATAAATGGTTTCAACGAAGATCGGGTGACAAAGCCATAGAGAAGATC 1020  
 QY 1069 AAATCTGCCAAGATAAATCGTCGAAGAGAGCTCGAGTCTTTTCAAGCCAACTGCC 1128  
 Db 1021 AAATCTGCCAAGATAAATCGTCGAAGAGAGCTCGAGTCTTTTCAAGCCAACTGCC 1080  
 QY 1129 ACCACATCAGCACCCCTAAAACGGAGAGAGCTTCGGATAAAACAAAGCAAGCAGCTCG 1188  
 Db 1081 ACCACATCAGCACCCCTAAAACGGAGAGAGCTTCGGATAAACAAAGCAAGCAGCTCG 1140  
 QY 1189 AACAGAATAAAGGCTGGTGGAAAGAGAAATAATCTTGGATGCTTGATGTACAACATA 1248  
 Db 1141 AACAGAATAAAGGCTGGTGGAAAGAGAAATAATCTTGGATGCTTGATGTACAACATA 1200  
 QY 1249 CGACTACGAAGAGCAGCGGTGGCGTGATCCTTCGCTTAGATATTTAACTCCCTGTTTTA 1308  
 Db 1201 CGACTACGAAGAGCAGCGGTGGCGTGATCCTTCGCTTAGATATTTAACTCCCTGTTTTA 1260  
 QY 1309 ACTCAGAGCTTTGGTAAAGTTTGCATGCTATGTTTCAAGCTGGGGTAAAGTTAGTTGTTG 1368  
 Db 1261 ACTCAGAGCTTTGGTAAAGTTTGCATGCTATGTTTCAAGCTGGGGTAAAGTTAGTTGTTG 1320  
 QY 1369 AAGAGATTGGTGTACCAAGTAAACAAACTTATCGCTGTTTTT 1411  
 Db 1321 AAGAGATTGGTGTACCAAGTAAACAAACTTATCGCTGTTTTT 1363

RESULT 5

AX02117  
 ID AX02117 standard; DNA; 1144 BP.  
 XX AX02117;  
 XX AC  
 XX XX  
 XX 23-APR-1999 (first entry)  
 XX DE Human FEN-1 DNA fragment.  
 XX KW FEN-1; human; flap endonuclease; detection; diagnosis; carcinogen;  
 XX KW neoplasia; antineoplastic agent; cleavage; ss.  
 XX OS Homo sapiens.  
 XX PN US5874283-A.  
 XX PD 23-FEB-1999.  
 XX PF 30-MAY-1995; 9505-0455968.  
 XX PR 30-MAY-1995; 9505-0455968.  
 XX PA (HARR/) HARRINGTON J J.  
 XX PA (HSIE/) HSIEH C.  
 XX PA (LIEB/) LIEBER M R.  
 XX PI Harrington JJ, Hsieh C, Lieber MR;  
 XX WP1; 1999-179985/15.  
 XX DR  
 XX XX  
 XX DNA encoding flap endonuclease polypeptides - useful for producing  
 XX PT e.g. recombinant polypeptides  
 XX PS Claim 3; Column 25; 58pp; English.  
 XX XX

CC This sequence encodes a human FEN-1 (flap endonuclease) protein. This  
 CC protein can be used in methods for detecting a pathological condition in  
 CC a patient, for diagnostic purposes, for screening for antineoplastic  
 CC agents and carcinogens, for diagnostic staging of neoplasia, for  
 CC producing recombinant flap endonuclease for use as research or  
 CC diagnostic reagents, for producing antibodies reactive with the novel  
 CC polypeptides, for producing transgenic nonhuman animals expressing the  
 CC novel polypeptides encoded by a transgene. The invention also provides  
 CC novel molecular cloning techniques and reagents involving cleavage of  
 CC a flap or nick with a flap endonuclease.  
 XX  
 SQ Sequence 1144 BP; 292 A; 283 C; 353 G; 216 T; 0 other;  
 Query Match 20.7%; Score 302.6; DB 20; Length 1144;  
 Best Local Similarity 56.6%; Pred. No. 1.6e-66;  
 Matches 601; Conservative 0; Mismatches 454; Indels 6; Gaps 2;  
 QY 85 ATGGGCATCAGGTTTTCAGCAAACTGCTGGCGGAAATGCGCCCAAGGCGGATGAAGAG 144  
 Db 1 ATGGGAATTCAGGCTGGCCAACTAATTCGTGTGTGGCCCGCCAGTCCCATCGGGAG 60  
 QY 145 CAGAAGTTCGAGAGCTACTTCGCGCGCAAAATCGCGTTCGAGCCGAGCATGAGCATATAC 204  
 Db 61 AATGACATCAAGAGCTACTTTGGCCGTAAGGTGGCCATTTGCTCTATGAGCATTTAT 120  
 QY 205 CAGTTCCTGATTGTTGAAGGACAGGCATGGAATACTCTCAAAATGAAGCTGGTGAA 264  
 Db 121 CAGTTCCTGATTGTTTCGCGAG--GGTGGGATGTGCTGCAGATGAGGAGGTGAG 177  
 QY 265 GTCACTAGTCAATTTGCAAGGAATGTTCAACCGGACATTAAGATTACTGGAAGCGGGAATC 324  
 Db 178 ACCACCGCCACCTGTGGGATGTTCTACCGCCACCATTCGCATGATGAGAGAGGCGCAT 237  
 QY 325 AAGCAGTTTATGTTTGTGCGCAAGCTCTCGATATGAAGAAACAAAGAGCTTGTCTAAA 384  
 Db 238 AAGCCGCTGATGCTTTGATGGCAAGCCGACAGCTCAAGTCAGGCGGAGCTGGCCANA 297  
 QY 385 AGATACCTCAAAAGAGATGATGCAACCAAGATCTGACTGAGCGAGTAGAGTAGAGAT 444  
 Db 298 CGCAGTGAAGCGGCGGTGAGGCAGAGAGCAGCTGAGCAGCGCTGCTGGGGCC 357  
 QY 445 AAGATCGGATGGAATAATTCAGCAAGAGGACTTAAAGGTCAAGGCAACACCAAGAA 504  
 Db 358 GAGCAGGAGGTGGAAATAATTCAGCGGCTGGTGAAGGTCACTAAGCAACCAATGAT 417  
 QY 505 GATTGTAAGCGCTATTAAAGACTTATGGGGTTCCTGTGTAGAGGACCTTCTGAAGCA 564  
 Db 418 GAGTGCAACATCTGCTGAGCCTCATGGCATCCCTTATCTTGATGCACCCAGTGAGCA 477  
 QY 565 GAGCAGATGTGAGCCCTTTGCAATAAAGATTAAGGTGTTGCTGTTGCTTCAAGAGAT 624  
 Db 478 GAGGCCAGCTGTGCTGCCCTGGTGAAGGCTGGCAAGTCTATGCTGGCGGTACCGAGGAC 537  
 QY 625 ATGAGCTCCCTTACTTTTGGGGCTCCACGCTTCCTCGTCATTTAATGGATCCAAAGTCC 684  
 Db 538 ATGAGCTCCCTCAGCTTCGCGGACCTGTGCTAATGCGACACCTGACTGCCAGTGAAGCC 597  
 QY 685 AAGAAAATACCTGTGATGGAATTGATGTTGCCAAGGTTTTGGAGGAGCTTGAAGTCAACC 744  
 Db 598 AAAAAAGCTGCCAATCCAGGAATCCACCTGAGCGGATTTCTGCAGGAGCTGGGCTGAAC 657  
 QY 745 ATGGACCAGTTTCAATTTGTCATCTCTGTGGATGTGACTATTGTGTAGTACATCAAA 804  
 Db 658 CAGGAACAGTTTGTGGATCTGTGATCTCTAGGCAAGTACTGTGTAGAGTATCCGG 717  
 QY 805 GGTATCGGGGGCAACAGAGCTCTGAAACTTATTCGTCAACATGGTCCATAGAAAGATC 864  
 Db 718 GGTATCGGGGGCAACAGCGGCTGTGGACCTCATCCAGNAGCAAGAGCATCGAGGAGATC 777  
 QY 865 TTGGAGATCTTAATAAGAGACAGATATCAAAATTCCTGAGGAGCTGGCTTTACCAAGAGCT 924  
 Db 778 GTGGCGGAGCTTGACCCCAACAAAGTACCCCTGTGCCAGAAAAATTTGGCTTCCCAAGAGGCT 837

QY 925 CGACGCTGTGTTCAAGGAGCCTAATG---TCACATTTGGATATTCCTGAGCTAAATGGACT 981  
 Db 838 CACCAGCTCTTCTTGGAACTGAGGTCTGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 897  
 QY 982 GCACCTGATGAGGAGGCTCATAAGTTTCTGTGTAAGAAAGATAATGTTTCAACGAAGAT 1041  
 Db 898 GAGCCAAATGAAGAAGAGCTGATCAAGTTCTATGTGTGTGTAAGAAAGCAGTTCTCTGAGGAG 957  
 QY 1042 CGGGTGACAAAGGCCATAGAGAAGATCAAAATCTCCCAAGATTAATCGTCCCAAGGAAGA 1101  
 Db 958 CGAATCCGCACTGGGTCAAGAGCTGAGTAAGAGCCGCCAAGGAGCACCAGGGCGGC 1017  
 QY 1102 CTCGAGTCCTTTTTCAGGCCAACTGCCACCACATCAGCACC 1142  
 Db 1018 CTGGATGATTCTTCAAGGTGACCGGCTCACTCTCTTCAGC 1058

RESULT 6

AA02107  
 ID AAX02107 standard; cdna; 1144 BP.

AC AAX02107;

XX 23-APR-1999 (first entry)

DE Human FEN-1 cDNA.

FW FEN-1; human; flap endonuclease; detection; diagnosis; carcinogen;  
 KW neoplasia; antineoplastic agent; cleavage; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 1..1144

FT CDS

FT /\*tag= a

FT /product= "FEN-1"

XX US5874283-A.

XX 23-FEB-1999.

XX 30-MAY-1995; 95US-0455968.

XX 30-MAY-1995; 95US-0455968.

XX (HARR/) HARRINGTON J J.

XX (HSIE/) HSIEH C.

XX (LIEB/) LIEBER M R.

XX Harrington JJ, Hsieh C, Lieber MR;

XX WPI: 1999-179985/15.

XX P-PSDB; AAW92504.

XX DNA encoding flap endonuclease polypeptides - useful for producing

XX e.g. recombinant polypeptides

XX Disclosure; Fig 1B; 58pp; English.

XX This sequence encodes a human FEN-1 (flap endonuclease) protein. This  
 CC protein can be used in methods for detecting a pathological condition in  
 CC a patient, for diagnostic purposes, for screening for antineoplastic  
 CC agents and carcinogens, for diagnostic staging of neoplasia, for  
 CC producing recombinant flap endonuclease for use as research or  
 CC diagnostic reagents, for producing antibodies reactive with the novel  
 CC polypeptides, for producing transgenic nonhuman animals expressing the  
 CC novel polypeptides encoded by a transgene. The invention also provides  
 CC novel molecular cloning techniques and reagents involving cleavage of  
 CC a flap or nick with a flap endonuclease.

XX Sequence 1144 BP; 292 A; 283 C; 353 G; 216 T; 0 other;

XX Query Match 20.7%; Score 302.6; DB 20; Length 1144;

Best Local Similarity 56.6%; Pred. No. 1.1e-66;  
 Matches 601; Conservative 0; Mismatches 454; Indels 6; Gaps 2;  
 QY 85 ATGGGATCAAGGGTTGACGAAACTGTGGCGGACAAATCGGCCCAAGGCGATGAAGGAG 144  
 Db 1 ATGGGAAATTCAGAGGCTGGCCAAACTAATTTGCTGATGTGGCCCGCAGTGCATCGGAG 60  
 QY 145 CAGAAGTTCCGAGAGCTACTTGGCCGCAAAATCGCGTCTGACGCCAGCATGAGCATATAC 204  
 Db 61 AATGACATCAAGAGCTACTTGGCCGTAAGTGGCCATGATGCCTCTATGAGCTTTAT 120  
 QY 205 CAGTTCTCTGATTTGTAAGGACAGGACATGGAACCTCTCACAAATGAAGCTTGGTAA 264  
 Db 121 CAGTTCTCTGATTTGTAAGGACATGGAACCTCTCACAAATGAAGCTTGGTAA 177  
 QY 265 GTCACCTAGTCATTTGCAAGGAATTTCAACCCGCAATAGATTTACTTGGAAAGCGGGAATC 324  
 Db 178 ACCACGACCCATGATGGCATTTCTACCGCACTTCGCATGATGGAAGACGGGATC 237  
 QY 325 AAGCCAGTTTATGTTTGTATGGCAAGCCTCTGATATGAAGAAACAAAGAGCTTGCATAA 384  
 Db 238 AAGCCGCTGATGTTTGTATGGCAAGCCTCTGATATGAAGAAACAAAGAGCTTGCATAA 297  
 QY 385 AGATACCTCAAAAGAGATGATGCAACCAAGATCTGACTGAGGAGTGTAGGATAGAGAT 444  
 Db 298 CGCAGTGAGCGGGCTGAGGCAGAGAAGCAGCTGAGCAGGCTCAGGCTGTGGGGCC 357  
 QY 445 AAGATGCGATTCAAAAATTTGACCAAGAGAGCTGTAAAGGTCACAGGCAACACAAAGAA 504  
 Db 358 GAGCAGGAGGTGGAATAATTTCACTAAGCGGCTGGTGAAGGTCTACTAAGCAGCAATGAT 417  
 QY 505 GATTGTAAACGGCTATTAAGACTTATGGGGTTCCTGTTGTAGAGGACCTTCTTGAAGCA 564  
 Db 418 GAGTGAACATCTGCTGAGCCTCATGGGCATCCCTATCTTGTATGACCCAGTGGCA 477  
 QY 565 GAAGCAGATGTGACCCCTTTTGCATAAAGTAAAGTGTTCGCTTGTGCTTCAAGAGAT 624  
 Db 478 GAGGCCAGCTGTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537  
 QY 625 ATGGACTCCCTTACTTTTGGGGTCCACGTTCTTCTGCTCATTTAATGGATCCAAAGTCC 684  
 Db 538 ATGGACTCCCTTACTTTTGGGGTCCACGTTCTTCTGCTCATTTAATGGATCCAAAGTCC 597  
 QY 685 AAGAAATACCTGCTGATGGAATTTGATTTGCCAAGGTTTTGGAGGAGCTTGAACCTACC 744  
 Db 598 AAGAGCTGCAATCCAGGAATTCACCTGAGCGGATTTCTGCAGGAGCTGGGCTGAC 657  
 QY 745 ATGGACCACTTCATTTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 804  
 Db 658 CAGGAACAGCTTTGTGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717  
 QY 805 GGTATCGGGGGGCAACAGCTCTGAAACTTATTCGTCAACATGGTCCATAGAAAGATC 864  
 Db 718 GGTATTTGGCCCAAGCGGGCTGTGGACCTCATCCAGAAGCAGCAGATCATGAGGAGAT 777  
 QY 865 TTGGAGAACTTTAATAAGACAGATATCAAAATTCCTGAGGAGCTGGCCTTACCAAGAGCT 924  
 Db 778 GTGGCGGACTTGACCCCAACAGTACCTCTGTCAGAGAAATTTGGCTCCACAGAGGCT 837  
 QY 925 CGACGCTTCTTCAAGGAGCTTAATG---TCACATTTGGATATTCCTGAGCTAAATGGACT 981  
 Db 838 CACCAGCTCTTCTTGGAACTGAGGTCTGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 897  
 QY 982 GCACCTGATGAGGAGGCTCATAAGTTTCTGTGTAAGAAAGATAATGTTTCAACGAAGAT 1041  
 Db 898 GAGCCAAATGAAGAAGAGCTGATCAAGTTCTATGTGTGTGTAAGAAAGCAGTTCTCTGAGGAG 957  
 QY 1042 CGGGTGACAAAGGCCATAGAGAAGATCAAAATCTCCCAAGATTAATCGTCCCAAGGAAGA 1101  
 Db 958 CGAATCCGCACTGGGTCAAGAGCTGAGTAAGAGCCGCCAAGGAGCACCAGGGCGGC 1017  
 QY 1102 CTCGAGTCCTTTTTCAGGCCAACTGCCACCACATCAGCACC 1142

Db 1018 CTGGATGATTTCTTCAAGGTGACCGGCTCACTCTCTTCAGC 1058

RESULT 7  
AA02111

ID AAX02111 standard; DNA; 2033 BP.

XX AC AAX02111;

XX DT 23-APR-1999 (first entry)

XX DE Human FEN-1 genomic DNA.

XX FEN-1; human; flap endonuclease; detection; diagnosis; carcinogen;  
KW neoplasia; antineoplastic agent; cleavage; ss.  
XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
FT CDS 104..1240  
FT /\*tag= a  
FT /product= "FEN-1"

XX PN US5874283-A.

XX PD 23-FEB-1999.

XX PF 30-MAY-1995; 95US-0455968.

XX PR 30-MAY-1995; 95US-0455968.

XX PA (HARR/) HARRINGTON J J.

XX PA (HSIE/) HSIEH C.

XX PA (LIEB/) LIEBER M R.

XX PI Harrington JJ, Hsieh C, Lieber MR;

XX DR WPI; 1999-179985/15.

XX DR P-PSDB; AAW92508.

XX PT DNA encoding flap endonuclease polypeptides - useful for producing  
XX PT e.g. recombinant polypeptides  
XX PS Disclosure; Fig 5A-B; 58pp; English.

XX CC This sequence encodes a human FEN-1 (flap endonuclease) protein. This  
CC protein can be used in methods for detecting a pathological condition in  
CC a patient, for diagnostic purposes, for screening for antineoplastic  
CC agents and carcinogens, for diagnostic staging of neoplasia, for  
CC producing recombinant flap endonuclease for use as research or  
CC diagnostic reagents, for producing antibodies reactive with the novel  
CC polypeptides, for producing transgenic nonhuman animals expressing the  
CC novel polypeptides encoded by a transgene. The invention also provides  
CC novel molecular cloning techniques and reagents involving cleavage of  
CC a flap or nick with a flap endonuclease.

XX SQ Sequence 2033 BP; 500 A; 493 G; 579 G; 461 T; 0 other;

Query Match 18.9%; Score 276.6; DB 20; Length 2033;  
Best Local Similarity 55.9%; Pred. No. 7.9e-50;  
Matches 638; Conservative 0; Mismatches 479; Indels 24; Gaps 5;

QY 85 ATGGGATCAAGGTTTTCAGCAAACTCTGGCGACAAATCGCCCAAGCGATGAAGAG 144

Db 104 ATGGAATTCACGGCTTGCACAACTAATTCCTGATGTGGCCCCCAGTCCCGTGAG 163

QY 145 CAGAAATTCAGAGCTACTTCGGCGAAATCGCCGTCAGCCACCATCAGCATATAC 204

Db 164 AATGACATCAAGAGCTACTTTGGTCGAAAGTGCCCATCGATCGCTCCATGAGCATCTAC 223

QY 205 CAGTCTCTGATGTAGTTGAAGACAGAGCATGGAACTCTCACAATGAAGCTGGTGA 264

Db 224 CAGTCTCTGATGTCTCTCAG---GGTGGGATGTGCTGCAGACAGAGGGGTGAG 280

QY 265 GTCCTAGTCAATTTGCAAGGAATTTCAACCGGACAATAAGATTACTTGGAGCGGGAATC 324  
Db 281 ACCACGAGCC---TGATGGGCATGTTCTACCGTACCATGG---CATGGAGAATGGCATC 334  
QY 325 AAGCCAGTTTATGTTTGGTGGCAAGCCCTCCTCATATGAAGAAACAAGAGCTTGTCTAAA 384  
Db 335 AAGCCTGTGACGTTCTTTTGGTGGCAAAACCCACAGCTGAAGTCAAGCGAGCTGCCAAG 394  
QY 385 AGATACTCAAAAAGAGATGATGCAACCAAGATCTGACTGAGGCAGTAGAGTAGAGAT 444  
Db 395 CGCAGTCAGAGGCGCGCGAGGCTGAGAACAATGAGCAGGCTCAGCAGGCTGGGATG 454  
QY 445 AAGATCGATTGAAAAATTGAGCAAGAGGACTGTAAGGTTCACAAGGCAACAACGAA 504  
Db 455 GAGGAGGAGGTGGAGAAGTTTCAACAAGAGGCTCGTAAGGTTCACAAGCAACAATGAT 514  
QY 505 GATTGTAACGGCTATTAAAGACTTATGGGGTTCTCTGTTGTAGAGGCACCTTCTGAAGCA 564  
Db 515 GAGTGCACAAACCTGCTGAGCCTCATGGGATCCCTTACCTTGATGCCAGCGAGGCA 574  
QY 565 GAAGCAGAATGTGCAGCCCTTTGCATAAAGATGAAGGTGTTCGCTGTTCTCAGAAGAT 624  
Db 575 GAGCCAGCTGTGCTGCCCTGGCAAAAGGCTGGCAAGTCTATGCTGCGGCCACGAGGAC 634  
QY 625 ATGACTCCCTTACTTTTGGGGTCCACGGTTCCTTCGTCATTTAATGGATFCCAAGTTCC 684  
Db 635 ATGACTGCCTCACTTTTGGCAGCCCGTCTAATGCGACACTTAAGTGCAGTGAGGCC 694  
QY 685 AAGAAAATACCTGTGATGGAATTTGATGTTGCCAAGGTTTGGAGGAGCTTGAACCTACC 744  
Db 695 AAGAAGCTGCCCATCCCAAGAGTTCCTATCTGAGCGGCTCTGACAGGAGCTGGGTCTGAAC 754  
QY 745 ATGACCAGTTCAATTTGTGTCATCCTGTGTGATGTGACTATTGTGATAGCATCAAA 804  
Db 755 CAGGAGCAGTTTGTGGATCTGTCATCCTGCTGGTAGCGACTACTCGGAGAGCATCGT 814  
QY 805 GGTATCGGGGGCAACAGCTCTGAAACTTATTCGTCAACATGGGTCCCTAAGAAGCATC 864  
Db 815 GGCATTGGCCCAAGCGGCTGTGGATCTCATCCAGAAACATAAGAGCATCGAGAGATC 874  
QY 865 TTGAGAATCTTAATAAGACAGATATCAAAATTCCTCAGGACTGGCCTTACCAGAAGACT 924  
Db 875 GTGAGCGGCTGGACCCCAAGTACCCCGTTCCAGAGAACTGGCTCCACAGGAAGCC 934  
QY 925 CGACGCTTGTTCAGGAGCGCTAA---TGTCACATTTGGATATTCCTGAGCTAAAATGGACT 981  
Db 935 CAGCAGCTCTTCTCTGGAGCCAGAGTAGTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 994  
QY 982 GCACCTGATGAGGAGGCTCATAAAGTTTCTGTTAAAGATATGTTTCAACGAGAT 1041  
Db 995 GAGCCAAATGAAGAAGAGTTGGTCAAAATTTATGTGTGTGAAAGAGCAGTTTTTTGAAGAG 1054  
QY 1042 CGGTGACAAAGGCCATPAGAGAAGATCAAAATCTGCCAAGAAATAAATCGTCGCAAGGAAGA 1101  
Db 1055 CGAATTCGAGTGGGTCAAGCGGCTGAGTAAGAGCCGCCAGGCGACACCCAGGGAGCC 1114  
QY 1102 CTCGAGTCTCTTTTTCAGGCCAACTGCCACCAACATCAGCACCCTGCTAAAACGAGGAGACT 1161  
Db 1115 CTCGATGATTTCTTCAAGGTGACAGGCTCACTCTCCTCAGC-----TAAGCGC 1162  
QY 1162 TCGGATAAACAAGCAAGGAGGCTGCGAACAAGAAACAAGAGCTGTGTGAAGAAGAAA 1221  
Db 1163 AAGGAGCAGAACCCAGGGGCTGCTAAGAAGAAAGCAAGAGCTGGGGGAGCGGGAAG 1222  
QY 1222 T 1222  
Db 1223 T 1223

RESULT 8  
ABL14287  
ID ABL14287 standard; cDNA; 1300 BP.



PR 30-MAY-1995; 95US-0455968.  
 XX (HARR/) HARRINGTON J J.  
 PA (HSIE/) HSIEH C.  
 PA (LIEB/) LIEBER M R.  
 XX  
 PI Harrington JJ, Hsieh C, Lieber MR;  
 XX WPI; 1999-179985/15.  
 DR P-PSDB; AAW92505.  
 XX  
 XX DNA encoding flap endonuclease polypeptides - useful for producing  
 PT e.g. recombinant polypeptides  
 XX  
 XX Disclosure; Fig 2B; 58pp; English.  
 XX  
 CC This sequence encodes a mouse FEN-1 (flap endonuclease) protein. This  
 CC protein can be used in methods for detecting a pathological condition in  
 CC a patient, for diagnostic purposes, for screening for antineoplastic  
 CC agents and carcinogens, for diagnostic staging of neoplasia, for  
 CC producing recombinant flap endonuclease for use as research or  
 CC diagnostic reagents, for producing antibodies reactive with the novel  
 CC polypeptides, for producing transgenic nonhuman animals expressing the  
 CC novel polypeptides encoded by a transgene. The invention also provides  
 CC novel molecular cloning techniques and reagents involving cleavage of  
 CC a flap or nick with a flap endonuclease.  
 XX  
 SQ Sequence 1930 BP; 482 A; 459 C; 553 G; 436 T; 0 other;  
 Query Match 18.6%; Score 272.2; DB 20; Length 1930;  
 Best Local Similarity 55.1%; Pred. No. 1e-58;  
 Matches 629; Conservative 0; Mismatches 488; Indels 24; Gaps 4;  
 QY 85 ATGGGATCAAGGTTTGAAGAACTCTGGCGGACAAATCGCCGAGCGATGAAGGAG 144  
 DB 1 ATGGGAATTCACGGCTTGGCAAACTAATGTCTGATGTGGCCGCCAGTCCCGTGAG 60  
 QY 145 CAGAAGTTCGAGAGTCTTCGCGCAAAATCGCGTGCACCGCAGCATGACATATAC 204  
 DB 61 AATGACATCAAGAGTCTTGTGTGAAGTGGCCATCGATCGCTCCCATGAGCATCTAC 120  
 QY 205 CAGTCTCTGATTTAGTGAAGCAGCAGCATGGAACTCTCACAATGAAGCTGGTGA 264  
 DB 121 CAGTCTCTGATTTCTGTCTGAG---GGTGGGATGTGCTGCAGAACAGGAGGTGAG 177  
 QY 265 GTCACTAGTCAATTTGAAGAAATGTTCAACCGGACAAATAGATTAATGGAAGCGGAATC 324  
 DB 178 ACCACGAGCTGTATGGGCTGTTAT-----GGCAAAACCATCCGATGGAGATGGCATC 231  
 QY 325 AAGCCAGTTTATGTTTGTATGCAAGCTCTCTGATATGAAGAAACAGAGCTTGTAA 384  
 DB 232 AAGCCGTGTACGTCTTGTATGCAACACCAACACAGCTGAAGTCAAGGAGCTGGCCAG 291  
 QY 385 AGATACTCAAAAGAGATGATGCAACCAAGATCTGACTGAGGCAGTAGAGTAGGAGAT 444  
 DB 292 CGCAGTGAGAGGCGCCGAGGCTGAGAAGCACTCAGCAGGCTCAGGAGGTGGATG 351  
 QY 445 AAGATGCGATGCAAAATTTAGCAGAGAGAGTGTAAAGTCAAGAGCAACACAGAA 504  
 DB 352 GAGGAGGAGGTGGAGAAGTTCACCAAGAGGCTGTGAAGTCAACCAAGCAACAAATAT 411  
 QY 505 GATTGTAACGGCTATTAAAGCTATTATGGGGTTCCCTGTTGTAGAGGCACCTTCTGAAGCA 564  
 DB 412 GAGTGAACACCTCTGTGAGCTCTATGGGCATCCCTTACCTTGATGCACCCAGCGAGCA 471  
 QY 565 GAAGCAGAAATGTGACGCCCTTTGCATTAACGATTAAGGTGTGCTGTTCTTCAGAAGAT 624  
 DB 472 GAGGCCAGCTGTGCTCCCTGGCAAGGCTGCAAAAGTCTATGCTGCGGCCAGGAGGAC 531  
 QY 625 ATGGACTCCCTTACTTTTGGGGCTCCACGGTCTCTTCGTCATTTAATGATCAAGTCC 684  
 DB 532 ATGGACTGCTCTACTTTTGGCAGCCCGCTGCTAATGCGACACTTAACTGCCAGTGAGGCC 591

QY 685 AAGAAAATACCTGTGTATGGAAATTTGATGTTGCCAAGGTTTGGAGGAGCTTGAACAC 744  
 DB 592 AAGAGCTGCCATCCCAAGAGTTCCATCTGAGCGGCTCTCGCAGGAGCTGGGTCTCAAC 651  
 QY 745 ATGACCAAGTTCAATTTGATTTGTCATCTGTGTGGATGTGACTATTGTGTAGATCAAA 804  
 DB 652 CAGGAGCAGTTTGTGGATCTGTGATCTCTGCTGGTAGCAGCTACTCGGAGAGCATCCGT 711  
 QY 805 GGTATCGGGGGCAACAGCTCTGAACATTTATTCGTCAACATGGTCCATAGAAGCATC 864  
 DB 712 GGATTTGGCCCAAGCGGGCTGTGGATCTCATCCAGAAACATTAAGAGCATCGAGGAGATC 771  
 QY 865 TTGGAGAATCTTAATAAGACAGATATCAAAATCTCTGAGGAGCTGGCTTACCAGAAAGCT 924  
 DB 772 GTGAGCGGCTGGACCCCAAGTACCCCGTTCAGAGAACTGGCTCCACAAGGAAGCC 831  
 QY 925 CGAGCTGTGTTCAAGAGAGCTAA---TGTCACATTTGATATTCCTGAGCTAAATGAGTACT 981  
 DB 832 CAGCAGCTCTTCTGGAGCCAGAGTAGTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 891  
 QY 982 GCACCTGATGAGGAGGCTCTATAAGTTTCTCTGTAAAGATATGTTTCAACGAAGAT 1041  
 DB 892 GAGCCAAATGAAGAGAGTGTGTCAAATTTATGTGGTGAAGAGAGTCTTCTGAAGAG 951  
 QY 1042 CGGTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAAATAAATCTGCGAAGGAAGA 1101  
 DB 952 CGAATTCGCAAGTGGGTCAAGCGGCTGAGTAAGAGCGCCAGGCGAGCACCCAGGGAGC 1011  
 QY 1102 CTCGAGTCTCTTTTCAAGCCAACTGCCACCATCATCAGCACCGCTAAACGGAAGGAGACT 1161  
 DB 1012 CTCGATGATTTCTTCAAGGTGACAGGCTCACTCTCTCAGC-----TAAGCGC 1059  
 QY 1162 TCGGATAAACAACAGCAGGCTGCGAACACAGAAACAAAGGCTGTGTGAAGAGAGAAA 1221  
 DB 1060 AAGGAGCCAGAACCCAGGGGCTGCTAAGAAGAAAGAAAGACTGGGGAGCGGGAG 1119  
 QY 1222 T 1222  
 DB 1120 T 1120  
 RESULT 10  
 AAX02109  
 ID AAX02109 standard; cdna; 1149 BP.  
 XX  
 XX AAX02109;  
 XX  
 DT 23-APR-1999 (first entry)  
 XX  
 DE Yeast FEN-1 cdna.  
 XX  
 KW FEN-1; yeast; flap endonuclease; detection; diagnosis; carcinogen;  
 KW neoplasia; antineoplastic agent; cleavage; ss.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN US5874283-A.  
 XX  
 PD 23-FEB-1999.  
 XX  
 PF 30-MAY-1995; 95US-0455968.  
 XX  
 PR 30-MAY-1995; 95US-0455968.  
 XX  
 PA (HARR/) HARRINGTON J J.  
 PA (HSIE/) HSIEH C.  
 XX (LIEB/) LIEBER M R.  
 PI Harrington JJ, Hsieh C, Lieber MR;  
 XX WPI; 1999-179985/15.  
 DR P-PSDB; AAW92506.



PT DNA encoding flap endonuclease polypeptides - useful for producing  
 PT e.g. recombinant polypeptides  
 XX  
 PS Disclosure; Fig 3B; 58pp; English.

XX This sequence encodes a yeast FEN-1 (flap endonuclease) protein. This  
 CC protein is used in a method to isolate novel human FEN-1 proteins for  
 CC detecting a pathological condition in a patient, for diagnostic purposes,  
 CC for screening for antineoplastic agents and carcinogens, for diagnostic  
 CC staging of neoplasia, for producing recombinant flap endonuclease for use  
 CC as research or diagnostic reagents, for producing antibodies reactive  
 CC with the novel polypeptides, for producing transgenic nonhuman animals  
 CC expressing the novel polypeptides encoded by a transgene. The invention  
 CC also provides novel molecular cloning techniques and reagents involving  
 CC cleavage of a flap or nick with a flap endonuclease.

XX Sequence 1149 BP; 410 A; 182 C; 279 G; 278 T; 0 other;  
 Query Match 18.5%; Score 270.2; DB 20; Length 1149;  
 Best Local Similarity 55.5%; Pred. No. 2.6e-58;  
 Matches 599; Conservative 0; Mismatches 453; Indels 27; Gaps 3;

QY 85 ATGCGGATCAAGGTTTGACGAACACTCTGGGACAAATCGGCCCAAGCGCATGAAGAG 144  
 DB 1 ATCGGTTATTAAGGTTTGAATGCAATATATATCGGAACATGTTCCCTCTCTATCAGGAA 60  
 QY 145 CAGAACTTCGAGAGCTACTTCGCGCGCAAAATCGCGTCGAGCGCAGCATGAGCATATAC 204  
 DB 61 AGCGATATCAAGAGCTTTTGGCAGAAAGTTGCCATCGATGCTCTATCTCTATAT 120  
 QY 205 CAGTTCCTGATTTAGTTGGAAGACAGGATGGAACACTCTCAAAATGAAGCTGGTGA 264  
 DB 121 CAGTTCCTGATTTAGTTGGAAGACAGGATGGAACACTCTCAAAATGAAGCTGGTGA 180  
 QY 265 GTCACTAGTCAATTCGAAAGATGTTCAACGGGACATATAGATTACTGGAAGCGGAAATC 324  
 DB 181 ACAACGTCACACTTGATGGGTATGTTTATAGGACACTGAGAATGATTGATAACGGTATC 240  
 QY 325 AAGCGATTTATGTTTGTGATGGAAGCTCTCTGATATGAAGAAACAGAGCTTGCTAAA 384  
 DB 241 AAGCGTTGTTATGTTTGTGATGGAAGCTCTCTGATATGAAGAAACAGAGCTTGCTAAA 300  
 QY 385 AGATACCTCAAAAGAGATGATGACCAACAGATCTGACTGAGCGAGTATAGAGTAGGAGAT 444  
 DB 301 CGGTCTCAAGAGAGGTGGAACAGAAACAACTGCGAGAGGCAACACAGAAATGGAA 360  
 QY 445 AAGATCGGATTTGAAATTTGAGCAAGAGGACTTAAAGTCAACAGGCAACACACAA 504  
 DB 361 AAGATGAAGCAAGAAAGA-----AGATTGTTGAAGGTTCTCAAAAGAGCATTAATGAA 411  
 QY 505 GATTGTAACGGCTATTAAAGACTTATGGGGTTCCTGTTGAGAGGACCTTCTGAAGCA 564  
 DB 412 GAAGCCCAAAATTTACTAGACTATGGAATCCATATATATAGCCCAACGGAAGCT 471  
 QY 565 GAAGCAGAAATGTCAGCCCTTTGTCATAAAGATAGGTGTTGCTGTTTCTCAGAAGAT 624  
 DB 472 GAGGCTCAATGTGCTGAGTTGGCAAGAAAGGAAAGGTGATGCGCGAGCAAGTGAAGAT 531  
 QY 625 ATGACTCCCTTACTTTTGGGCTCCAGGTTCTCTGCTCATTTAATGGATCCAGTTCC 684  
 DB 532 ATGACACACTCTGTTTATAGAACACCTTCTGTTGAGACATTTGACTTTTTCAGAGGCC 591  
 QY 685 AAGAAAATACCTGTGATGGAATTTGATGTTGCCAAGGTTTGGAGGAGCTTGAACCTACC 744  
 DB 592 AAGNAGAACCGATTACGAAATAGACTGAATAGTTTGGAGGAGCTGACTTGACA 651  
 QY 745 ATGACACAGCTCAATGATTTGTCATCTGCTGATGTTGATGTTGATGATGATGATGATGAT 804  
 DB 652 ATAGAGCAGTTGTTGATCTTTGTCATAATGCTTGGTCTGACTACTGTAAGCATCAGA 711  
 QY 805 GGTATCGGGGGGCAACAGCTCTCAAACTTATTCGTCAACATGGGTCCATAGAAAGCATC 864  
 DB 712 GGTGTTGGTCCAGTCGACAGCGCTTAAATTTGATAAAAGCGCATGGATCCATCGAAAAATC 771

QY 865 TTGAGAAATCTT-----AATAAAGACAGATATCAAAATTCCTGAGGACTGG 909  
 DB 772 GTGGAGTTTATTGAATCTGGGAGTCAACACACATTAATGGAATCCCAAGAGACTGG 831  
 QY 910 CTTTACCAGAGCTGCGAGCTTTGTTCAAGGAGCCCTTAATGTCACATTGGATA---TTCCT 966  
 DB 832 CTTTACAAAGCAAGAAATGCTGTTTCTTGACCTGGAAGTTATAGATGCTAAGCAATA 891  
 QY 967 GAGCTAAATGCACTGCACCTGATGAGGAGGCTCTCATAGTTTCTCTGGTAAAGATAAT 1026  
 DB 892 AACTTGAATGTGCGCCACCAAGGAGAGAACTTATCGAGTATTATGTGATGATAAG 951  
 QY 1027 GCTTTCAACAGAGATCGGTGACAAAGGCGCATAGAGAAGATCAAACTGCAAGAAATAA 1086  
 DB 952 AATTCAGTGAAGAAGAGTTAAATCTGTTATATCAAGATTGAAGAAGCTTGAAATCT 1011  
 QY 1087 TCGTCGCAAGAGACTCGAGTCTCTTTTCAAGCCAACTGCCACCATCAGACCCGCT 1145  
 DB 1012 GGCATTGAGGTAGTTAGATGGTCTTCTTCAAGTGTGCTTAAGACAAAGAACAGCT 1070

RESULT 11

ABL14286  
 ID ABL14286 standard; cDNA; 3362 BP.  
 XX  
 AC ABL14286;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 37340.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers BW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB70183.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX interactions -  
 PS Claim 1; SEQ ID NO 37340; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3362 BP; 979 A; 709 C; 750 G; 924 T; 0 other;  
 SQ

[illegible]

Db 244 GAGGACCAAGATGCCCGTCAAGAGTTCACTAGGACAAGCTGTGTGAAGTCTGGCC 303  
 Qy 739 CTCACATGGACCAAGTTCATTGATTTGGTCATCCCTGTGTGGATGTGACATTTGTGATAGC 798  
 Db 304 ATTAACAATCGAGAGTTCATTGATCTATGTATCTGCTGGTTCGATTTACTGTGAGAGC 363  
 Qy 799 ATCAAGGTATCGGGGGCAACAGCTCTGAACTCTGAACTATTTCTGTCACATGGTCCATAGAA 858  
 Db 364 ATCAAGGTATTTGGACCAAGCGGATCGAAGTATCAACACCTATCGGGATATAGAG 423  
 Qy 859 AGCATCTTTGGAGATCTTAATAAGACAGATATCAAAATTCCTGAGGACTGGCCTTACCAA 918  
 Db 424 ACTATCTTGGATAACCTGGACTCTAGCAATACACCGTCCCGGAGAACTGGAACATAAG 483  
 Qy 919 GAAGTCGACGGTCTGTCAGGAGCTTAATGTCACATTTGATATTCCT---GAGCTAAAA 975  
 Db 484 GTGGCGCGGAACCTTTCATCGAACCGGAGGTAGCTGATGCGGACTCCATAGATCTCAA 543  
 Qy 976 TGGACTGCACCTGATGAGGAGGTCTCATAGTTTCTGTTAAAGATATGTTTCAAC 1035  
 Db 544 TGGTTCGAGCCGATGAGGAGGCTTGTCAAGTTCTCTGCGGCGACCGGAGTTCAAC 603  
 Qy 1036 GAAGATCGGGTGACAAAGCCATAGAGAAGATCAAAATCTGCCAAGAAATAAATCGTCGAA 1095  
 Db 604 GAAGAGCGGTTTCGCAACGGTCCAAAAGCTGATGAATCCAAAGCAGGCCAGACTCAG 663  
 Qy 1096 GGAAGACTCGAGTCTCTTTTCAAGCCAACTGCCACCAC 1133  
 Db 664 GTGAGACTCGATAGCTTCTTTAAGACACTGCCCAGCAC 701

RESULT 13  
 AAT76685  
 ID AAT76685 standard; DNA; 1023 BP.  
 XX AC AAT76685;  
 XX DT 14-APR-1998 (first entry)  
 XX DE Pyrococcus furiosus FEN-1 endonuclease gene coding sequence.  
 XX KW Nucleic acid cleavage; 5' nuclease; DNA cleavage; RNA cleavage;  
 XX KW invader directed cleavage; FEN-1; endonuclease; ds.  
 XX OS Pyrococcus furiosus.  
 XX PN WO9727214-Al.  
 XX PF 31-JUL-1997.  
 XX PR 22-JAN-1997; 97WO-US01072.  
 XX PR 02-DEC-1996; 96US-0759038.  
 XX PR 24-JAN-1996; 96US-0599491.  
 XX PR 12-JUL-1996; 96US-0682853.  
 XX PR 29-NOV-1996; 96US-0756386.  
 XX PR 02-DEC-1996; 96US-0758314.  
 XX PA (THIR-) THIRD WAVE TECHNOLOGIES INC.  
 XX PI Brow MAD, Dahlberg JE, Hall JG, Kaiser MW, Lyamichev VI;  
 XX PI Olive DM, Prudent JR;  
 XX XX  
 XX WPI: 1997-393613/36.  
 XX P-PSDB; AAW24216.  
 XX PT Thermostable structure-specific nuclease(s) - used for detection and  
 XX PT characterisation of nucleic acid sequences and variations in nucleic  
 XX PT acid sequences  
 XX PS Example 28; Page 283-285; 457pp; English.  
 XX XX  
 XX CC This sequence comprises the coding region of the gene encoding

CC Pyrococcus furiosus (Pfu) FEN-1 endonuclease (see AAW24216).  
 CC It was obtained by PCR amplification (see AAT76682-83). Large  
 CC scale preparation of recombinant Pfu FEN-1 was performed using  
 CC E. coli as host. Pfu FEN-1 is a thermostable enzyme. It can be  
 CC used in novel methods for the detection and characterisation of  
 CC nucleic acid sequences and variations in nucleic acid sequences.  
 XX SQ Sequence 1023 BP; 396 A; 172 C; 240 G; 215 T; 0 other;  
 Query Match 10.9%; Score 159.6; DB 18; Length 1023;  
 Best Local Similarity 54.6%; Pred. No. 2.5e-30;  
 Matches 318; Conservative 0; Mismatches 264; Indels 0; Gaps 0;  
 Qy 134 CGATGAAGGACGAAGTTCGAGAGCTACTTCGGCGCAAAATCGCGTCGACGCAGCA 193  
 Db 29 CAAGAAAGAATTTAGATTAGAAACCTATACGGGAAAAAATCGCATCGAGCTCTTA 88  
 Qy 194 TGACATATACCACTCTCTGATTTGAGTTGGAAGGACGAGCATGAACTCTCAAAATG 253  
 Db 89 ATGCAATCTACCAATTTTGTCCACAATAAGACAGAAAAGATGGAATCCACTTATGATT 148  
 Qy 254 AAGCTGGTGAAGTCACTAGTCAATTTGCAAGGAATCTTCAACCGGACAATAAGATTACTGG 313  
 Db 149 CAAAGGGTAGAATAACCTCCACCTTAAGCGGCTCTTTTACAGGACAATAAACCTTAATGG 208  
 Qy 314 AAGCGGAATCAAGCCAGTCTTATGTTTGTGCAAGCTCTCTGATATGAAGAAACAAG 373  
 Db 209 AGGCTGGATATAACCTGTGTATGTTTGTGAGAGAACTCCAGAAATCAAAAAGAAAG 268  
 Qy 374 AGCTTGTCTAAAGATACCTCAAAAAGAGATGATGCAACCAAGATCTGACTGAGCGAGTAG 433  
 Db 269 AGCTCGAAAAAAG 328  
 Qy 434 AGGTAGGAGATAGATCGGATTTGAAAATTTGAGCAAGAGAGACTGTAAAGGTCAACAAGGC 493  
 Db 329 AAAAAGGAGATAGAGAGAGCAAGAAATATGCCCAAGAGCAACCGGTTAAATGAAA 388  
 Qy 494 AACACACGAAGATTTAAACGGCTATTAAAGCTTTATGGGGTTCCTGTTGTAGAGGCAC 553  
 Db 389 TGCTCATCGAGGATGCAAAAAAATCTTAGAGCTTATGGGAATTCCTATAGTTCAAGCAC 448  
 Qy 554 CTTCGAG 613  
 Db 449 CTAGCGAGGAG 508  
 Qy 614 CTTCGAGAGATAG 673  
 Db 509 CTAGTCAAGATTTACGATTTCCCTACTCTTTTGGAGCTCCAAGACTTGTAGAACTTAACA 568  
 Qy 674 ATCCAAGTTCCAAGAAATACCTGTGATGGAATTTGATGTTG 715  
 Db 569 TAACAGGAAAAAGAAAGTTGCTCGGAAAAAATGCTACGTCG 610  
 RESULT 14  
 AAV65840  
 ID AAV65840 standard; DNA; 1023 BP.  
 XX AC AAV65840;  
 XX DT 02-FEB-1999 (first entry)  
 XX DE Pyrococcus furiosus FEN-1 endonuclease gene ORF.  
 XX KW Nucleic acid detection; multiple sequential invasive cleavage;  
 XX KW FEN-1; endonuclease; nuclease; ds.  
 XX OS Pyrococcus furiosus.  
 XX PN WO9842873-Al.  
 XX XX  
 XX PD 01-OCT-1998.

PF	24-MAR-1998;	98WO-USO5809.
XX		
PR	24-MAR-1997;	97US-0823516.
XX		
PA	(THIR-) THIRD WAVE TECHNOLOGIES INC.	
XX		
PI	BROW MAD, Hall JG, Kwiatkowski RW, Lyamichev VI;	
PI	Mast AL, Vavra SH;	
XX		
DR	WPI; 1998-557036/47.	
XX	P-PSDB; AAW79970.	
DR		
XX	Detecting target nucleic acid by sequence-specific cleavage of	
PT	complex with two specific oligonucleotides - used to detect	
PT	Cytomegalovirus DNA	
XX		
PS	Example 28b; Page 316-317; 524pp; English.	
XX		
CC	This is the nucleotide sequence of the open reading frame encoding	
CC	the FEN-1 endonuclease (see AAW79970) of Pyrococcus furiosus (Pfu).	
CC	It was obtained from genomic DNA by PCR amplification (see	
CC	also AAU65838-39). The PCR product has been ligated into vector	
CC	pTrc99a, and FEN-1 was expressed in E. coli cells. The invention	
CC	relates to means for the detection and characterisation of nucleic	
CC	acid sequences, and variations in nucleic acid sequences. It also	
CC	relates to methods for forming a nucleic acid cleavage structure on	
CC	a target sequence and cleaving this structure in a site-specific	
CC	manner, preferably using a thermostable structure-specific nuclease	
CC	such as FEN-1. Cleavage of the cleavage structure by the	
CC	nuclease indicates the presence of specific nucleic acid sequences	
CC	or specific variants. The invention further relates to methods for	
CC	the separation of nucleic acid molecules based on charge, methods	
CC	for the detection of non-target cleavage products via the formation	
CC	of a complete and activated protein binding region, and methods for	
CC	the detection of nucleic acid from various viruses (e.g. human	
CC	Cytomegalovirus) in a sample. The method amplifies the detection	
CC	molecule rather than the target itself, is less subject to	
CC	contamination than exponential amplification processes, and allows	
CC	many targets to be analysed in a single reaction.	
XX		
SQ	Sequence 1023 BP; 396 A; 172 C; 240 G; 215 T; 0 other;	
	Query Match	10.9%; Score 159.6; DB 19; Length 1023;
	Best Local Similarity	54.6%; Pred. No. 2.5e-30;
	Matches 318; Conservative	0; Mismatches 264; Indels 0; Gaps
OY	134 CGATGAAGCAGCAGAAGTTTCGAGAGCTACTTCGGCGCAAAATCGCGTCGACGCCAGCA 193	
Db	29 CAAGAAAAGAATTGAGTTAGAAAACCTATACGGGAAAAAATCGCAANTCGAGCTCTTA 88	
OY	194 TGAGCATATACCAGTTCCTGTAGTTGATGGGAAGCAGGCATGGAAACTCTCACAAATG 253	
Db	89 ATGCAATCTACCANAATTTTTGTCCCAATAAGACAGAAAGATGGAATCCCACTTATG 148	
OY	254 AAGCTGGTGAAGTCACTAGTCTCATTTTCAAGGAATGTTCAACCGGACAATAAGTACTCG 313	
Db	149 CAAAGGGTAGATATACCTCCCACCTAACGGGCTCTTTTACAGGACATAAACCTAATGG 208	
OY	314 AAGCGGGAAATCAAGCCAGTTTATGTTTTTGATGGCAAGCCTCCTGATATGAAGAACAAG 373	
Db	209 AGCGCTGGATAAAGACCTGTGTATGTTTTTGTATGGAGAACCTCCAGAAATTCAAAAAGAA 268	
OY	374 AGCTTGTCTAAAAGATACTCAAAAAGCAGATGATGCAACCAAGCATCTGACTGAGGCAGTAG 433	
Db	269 AGCTCGAAAAAAGAGAGAACGGCAGAGGAAGAGCTGGAAGAAAAGTGGAGAGAACACTTG 328	
OY	434 AGGTGAGAGATAAAGATGCGATTTGAAAATTCAGCAACAGGACTGTAAGAGGTCAACAAGC 493	
Db	329 AAAAAGGAGAGATACAGGAAGCAAGAAAATATGCCCCAAGAGCAACACCGGTAAATGAA 388	
OY	494 RACACAAAGCAGATTTGTAACGGCTTATTAAGACTTATGGGGTTCCTGTGTGAGAGGCAC 553	
Db	389 TGCTCATCGAGATGCAAAAACACTCTTAGAGCTTATGGAAATTCCTAGTTCGAAGC 448	

```

QY 134 CGATGAGGAGCAGAACTGCGAGAGCTACTTGGCCGCAAAATCGCGTCGCGCCAGCA 193
Db 29 CAAGAAAGAAATTTGAGTTAGAAACCTATACGGGAAAAAATCGCATCGCGCTCTTA 88
QY 194 TGAGCATATACCGATTCCTGATTTAGTTGGAAGGACAGGCATGGAACCTCTCACAATG 253
Db 89 ATGCAATCTACCAATTTTGTCCACAATAAGACAGAAAGATGGAACCTCCACTTATGGATT 148
QY 254 AAGCTGGTGAAGTCACATAGTTCGAAGGAATTTCAACCGGACACATAGATTACTGG 313
Db 149 CAAGGGTAGAATAACCTCCACTAGCGGGCTCTTTACAGGACATAAACCTAATGG 208
QY 314 AAGCGGGAATCAAGCCAGTTTATGTTTTTGGTCAAGCCCTCCTGATATGAAGAAACAAG 373
Db 209 AGGCTGGAATAAAACCTGTGTATGTTTTGATGGAGAACCTCCAGAAATTCAAAAGAAAG 268
QY 374 AGCTTGTCTAAAGATCTCAAAAGAGATGATGCAACCAAGATCTGACTGAGGACAGTAG 433
Db 269 AGCTCGAAAAAAGAGAGAGAGAGGAGGAGCTGAAGAAAGTGGAGAGAGACACTTG 328
QY 434 AGGTAGGAGATAAAGATGCGGATTGAAAAATTTAGCAAGAGGACTGTAAAGCTCACAGGC 493
Db 329 AAAAGGAGAGATAGAGGAAGCAAGAAATATGCCCAAGAGAACCCAGGGTAAATGAAA 388
QY 494 AACACACGAAGATTGTAAACGGCTATTAAAGACTTATGGGGTTCTCTGTTAGAGGCAC 553
Db 389 TGCTCATCGAGGATGCAAAAAAACTTTAGAGCTTATGGGAATTCCTATAGTTCAAGCAC 448
QY 554 CTTCTGAAGCAGACAGAGATGTGCAGCCCTTTGCAATAAAGGATAAGGTGTCGCTGTTG 613
Db 449 CTAGCGAGGAGAGAGGCCCAAGCTGCATATATGGCCGCAAGGGGAGCGGTGATGCATCGG 508
QY 614 CTTCAAGAAGATATGGACTCCCTTACTTTTGGGGCTCCACGCTTCTTCGTCATTAAATGG 673
Db 509 CTAGTCAAGATTACGATTCCTACTTTTGGAGCTCCAGACTTGTAGAACTTAACAA 568
QY 674 ATCCAAGTTCCAGAAAAATACCTGTGATGGAATTTGATGTTG 715
Db 569 TAACAGGAAAAAGAAAGTTGCTGGGAAAAATGTCTACGTGG 610

```

Search completed: May 8, 2003, 19:58:37  
 Job time : 371 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 19:50:47 ; Search time 95 Seconds  
(without alignments)  
4722.822 Million cell updates/sec

Title: US-09-805-311-1

Perfect score: 1463  
Sequence: 1 caccgagaatagctcgccgccc.....aaaaaaaaaaaaaaaaaaaaa 1463

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2.6/ptodata/2/ina/PCUTUS\_COMB.seq.\*  
6: /cgn2.6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1463	100.0	1463	4	US-09-426-557-1
2	1421.8	97.2	1478	4	US-09-426-557-7
3	1409.8	96.4	1541	4	US-09-426-557-3
4	1350.2	92.3	1381	4	US-09-426-557-5
5	302.6	20.7	1144	2	US-08-455-968E-2
6	302.6	20.7	1144	2	US-08-455-968E-28
7	276.6	18.9	2033	2	US-08-455-968E-9
8	272.2	18.6	1930	2	US-08-455-968E-4
9	270.2	18.5	1149	2	US-08-455-968E-6
10	159.6	10.9	1023	2	US-08-757-653-175
11	159.6	10.9	1023	2	US-08-823-516-78
12	159.6	10.9	1023	3	US-08-759-038-114
13	159.6	10.9	1023	3	US-08-758-314-114
14	135.6	9.3	1032	4	US-09-146-319-1
15	135.6	9.3	1032	4	US-09-175-973-1
16	89.2	6.1	981	2	US-08-757-653-171
17	89.2	6.1	981	2	US-08-823-516-74
18	89.2	6.1	981	3	US-08-759-038-110
19	89.2	6.1	981	3	US-08-758-314-110
20	65.2	4.5	7218	1	US-08-232-463-14
21	47	3.2	455	1	US-08-636-928-4
22	47	3.2	1736	3	US-09-182-816-22
23	47	3.2	1736	3	US-09-182-816-24
24	47	3.2	1736	3	US-09-471-528-22
25	47	3.2	1736	3	US-09-471-528-24
26	47	3.2	1736	4	US-09-634-530-22
27	47	3.2	1736	4	US-09-634-530-24

```

28 46.8 3.2 2196 4 US-09-149-476-163 Sequence 163, Appl
29 46.4 3.2 685 4 US-09-227-357-66 Sequence 66, Appl
30 45 3.1 3246 3 US-09-005-180A-2 Sequence 2, Appl
31 45.4 3.1 1359 4 US-09-387-574-11 Sequence 11, Appl
32 45.4 3.1 1359 4 US-09-668-096-11 Sequence 11, Appl
33 45.2 3.1 873 4 US-09-475-316A-20 Sequence 20, Appl
34 45.2 3.1 2205 3 US-08-888-077A-41 Sequence 41, Appl
35 45 3.1 222 4 US-08-481-190-15 Sequence 15, Appl
36 45 3.1 222 5 PCT-US93-00869-15 Sequence 15, Appl
37 44.4 3.0 1159 4 US-09-410-464-14 Sequence 14, Appl
38 44.4 3.0 3410 4 US-09-020-956-110 Sequence 110, App
39 44.4 3.0 3410 4 US-09-030-607-110 Sequence 110, App
40 44.4 3.0 3410 4 US-09-605-785-110 Sequence 110, App
41 44.4 3.0 3410 4 US-09-439-313-110 Sequence 110, App
42 44.4 3.0 3410 4 US-09-352-616A-110 Sequence 110, App
43 44.4 3.0 3410 4 US-09-602-877A-100 Sequence 100, App
44 44.4 3.0 3410 4 US-09-232-149A-110 Sequence 110, App
45 43.8 3.0 1223 3 US-09-154-874-4 Sequence 4, Appl

```

#### ALIGNMENTS

```

RESULT 1
US-09-426-557-1
; Sequence 1, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
; FILE REFERENCE: 0961
; CURRENT APPLICATION NUMBER: US/09/426,557
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/112,332
; EARLIER FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1463
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (85)...(1221)
US-09-426-557-1

```

```

Query Match 100.0%; Score 1463; DB 4; Length 1463;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CACGAGATAGCTCGCGCGCGCGGTTCTTTCGCGCACTCCGGCTCAGCGCGCGCGCCA 60
Db 1 CACGAGATAGCTCGCGCGCGCGGTTCTTTCGCGCACTCCGGCTCAGCGCGCGCGCCA 60
Qy 61 CCCGCCACAGCCGCGAGACGAGATGGCATCAAGGTTTGACGAACTGCTGCGGAC 120
Db 61 CCCGCCACAGCCGCGAGACGAGATGGCATCAAGGTTTGACGAACTGCTGCGGAC 120
Qy 121 AATCGGCCAAGGCGATGAAGGACAGAAAGTTCGAGAGCTACTTCGCGCGCAATCGCC 180
Db 121 AATCGGCCAAGGCGATGAAGGACAGAAAGTTCGAGAGCTACTTCGCGCGCAATCGCC 180
Qy 181 GTCGAGCCACATGACATATACCTCTGATTTAGTTGGAAGGACAGGATCGAA 240
Db 181 GTCGAGCCACATGACATATACCTCTGATTTAGTTGGAAGGACAGGATCGAA 240
Qy 241 ACTCTCAAAATGAAGCTGGTGAAGTCACTAGTCACTTTCGAAGGAATGTTCAACCGGACA 300
Db 241 ACTCTCAAAATGAAGCTGGTGAAGTCACTAGTCACTTTCGAAGGAATGTTCAACCGGACA 300
Qy 301 ATAGATTACTGGAAGCGGGAATCAAGCCAGTTATGTTTGTATGCGCAATSCCTCTGAT 360

```







QY	859	AGCATCTTGGAGAAATCTTAATAAAGACAGATATCAAAATTCCTGAGACTGGCCTTACCAA	918
Db	853	AGCATCTTGGAGAAATCTTAATAAAGACAGATATCAAAATTCCTGAGACTGGCCTTACCAA	912
QY	919	GAAGCTGAGCGCTTGTTC AAGGAGCCTAATGTACATATGGATATTCCCTGAGCTAAATATGG	978
Db	913	GAAGCTGAGCGCTTGTTC AAGGAGCCTAATGTACATATGGATATTCCCTGAGCTAAATATGG	972
QY	979	ACTGCACCTGATGAGGAGGCTCTATAAGTTCCTCGTAAAAAGATAATGGTTTCAACGAA	1038
Db	973	ACTGCACCTGATGAGGAGGCTCTATAAGTTCCTCGTAAAAAGATAATGGTTTCAACGAA	1032
QY	1039	GATCGGTTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGATAAATCTGTCGAAGGA	1098
Db	1033	GATCGGTTGAGAAGGCCATAGAGAAGATCAAAATCTGCCAAGATAAATCTGTCGAAGGA	1092
QY	1099	AGACTCGAGTCCCTTTTTTCAAGCCAACTGCCACCACATCAGCACCGCTAAAACGGAAGGAG	1158
Db	1093	AGACTCGAGTCCCTTTTTTCAAGCCAACTGCCACCACATCAGCACCGCTAAAACGGAAGGAG	1152
QY	1159	ACTTCGGATAAAAAACAAGCAGCTGCGAACAAGAAAAACAAGGCTTGGTGGAAAGAAG	1218
Db	1153	ACTTCGGATAAAAAACAAGCAGCTGCGAACAAGAAAAACAAGGCTTGGTGGAAGAAG	1212
QY	1219	AAATAAATCTTGGATGCTTGATGTACAACATACGACATACGAAAGACAGCGGTGGCGTGATCAC	1278
Db	1213	AAATAAATCTTGGATGCTTGATGTACAACATACGACATACGAAAGACAGCGGTGGCGTGATCAC	1272
QY	1279	TTTCGCTTAGATATATTTAACTCCCTGTTTTAACTCAGAGCTTTGGTAAAAGTTTGCCTCATG	1338
Db	1273	TTTCGCTTAGATATATTTAACTCCCTGTTTTAACTCAGAGCTTTGGTAAAAGTTTGCCTCATG	1332
QY	1339	TTTCAAGCTGGGTAAAGTTAGTTGTGTTTGAAGAGATTGGTGTACCAAGTAACAAAACTT	1398
Db	1333	TTTCAAGCTGGGTAAAGTTAGTTGTGTTTGAAGAGATTGGTGTACCAAGTAACAAAACTT	1392
QY	1399	ATCGCTGTTTTTTTACTCTCTTTGCTTTGAAGTA	1431
Db	1393	ATCGCTGTTTTTTTACTCTCTTTGCTTTGAAGTA	1425

## RESULT 4

```

RES001 4
US-09-426-557-5
; Sequence 5, Application US/09426557
; Patent No. 623257
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0961
; CURRENT APPLICATION NUMBER: US/09/426,557
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/112,332
; EARLIER FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37)...(1173)
US-09-426-557-5

```

Query Match 92.3%; Score 1350.2; DB 4; Length 1381;  
Best Local Similarity 99.4%; Pred. NO. 0;  
Matches 1355; Conservative 0; Mismatches 8; Indels 0;

**Qy** 49 CGCGCCGCCACCGCCACAGCCGCGGCATCAGGGTTGACGAA 108  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
**Db** 1 CGACCCACGGTCGCGCCACAGCCGCGCATCAGGGTTGACGAA 60

QY	109	CTGTGGCGGACAATGCGCCCAAGCGGATGAAGGAGCAGAAGTTCGAGAGCTACTTTCGCG	166
DB	61	CTGTCTGGGCGGCAAAATGCGCCCAAGCGGATGAAGGAGCAGAAGTTCGAGAGCTACTTTCGCG	120
QY	169	CGCAAAATCGCGCTCGAGCCAGCATGAGCATATACCAGTTCTGATGTGTAGTTGGAAGG	228
DB	121	CGCAAAATCGCGCTCGAGCCAGCATGAGCATATACCAGTTCTGATGTGTAGTTGGAAGG	180
QY	229	ACAGCATGGAAACTCTCACAAATGAAGCTGGTGAAGTCACTAGTCAATTTGCAAGGAATG	288
DB	181	ACAGCATGGAAACTCTCACAAATGAAGCTGGTGAAGTCACTAGTCAATTTGCAAGGAATG	240
QY	289	TTCAACCGGCAAAATAAGATTACTCGAAGCGGGATCAAGCCAGTTTATGTCTTTTGTATGC	348
DB	241	TTCAACCGGCAAAATAAGATTACTCGAAGCGGGATCAAGCCAGTTTATGTCTTTTGTATGC	300
QY	349	AAGCCTCCTGATATGAAGAAACAAGACTTTGCTAAAAGATACTCAAAAAGAGATGATGCA	408
DB	301	AAGCCTCCTGATATGAAGAAACAAGACTTTGCTAAAAGATACTCAAAAAGAGATGATGCA	360
QY	409	ACCAAGATCTGACTGAGGCAGTAGAGGTAGGAGATAAAGATCGGATTCAAAAATTGAGC	468
DB	361	ACCAAGATCTGACTGAGGCAGTAGAGGTAGGAGATAAAGATCGGATTCAAAAATTGAGC	420
QY	469	AAGAGGACTGTAAAGGTCACAAGGCAACAACGAAAGATTGTAAACGGCTATTAAGACTT	528
DB	421	AAGAGGACTGTAAAGGTCACAAGGCAACAACGAAAGATTGTAAACGGCTATTAAGACTT	480
QY	529	ATGGGGGTTCTGTTGTAGAGCACTTTCTGAAGCAGAAGCAGAAATGTCAGCCCTTTGC	588
DB	481	ATGGGGGTTCTGTTGTAGAGCACTTTCTGAAGCAGAAGCAGAAATGTCAGCCCTTTGC	540
QY	589	ATAAACGATAAGGTGTTGCGTGTGCTTCAGAAGATATGGACTCCCTTACTTTTGGGCT	648
DB	541	ATAAACGATAAGGTGTTGCGTGTGCTTCAGAAGATATGGACTCCCTTACTTTTGGGCT	600
QY	649	CAOAGGTTCTCTCATTTAATGGATCCAAGTTCCAAGAAATACCTGTGATGAAATTT	708
DB	601	CAOAGGTTCTCTCATTTAATGGATCCAAGTTCCAAGAAATACCTGTGATGAAATTT	660
QY	709	GATGTTGCCAAGGTTTGGAGGAGCTTGAACTCACCATGGACAGTTCAATGATTTGTC	768
DB	661	GATGTTGCCAAGGTTTGGAGGAGCTTGAACTCACCATGGACAGTTCAATGATTTGTC	720
QY	769	ATCCTGTGTGGATGTAATTTGTGATAGCATCAAAAGTATCGGGGGCAACAGCTCTG	828
DB	721	ATCCTGTGTGGATGTAATTTGTGATAGCATCAAAAGTATCGGGGGCAACAGCTCTG	780
QY	829	AAACTTATTGTCACATGGGTCATAGAAAGCATCTTGGAGAACTTAAATAAGACAGA	888
DB	781	AAACTTATTGTCACATGGGTCATAGAAAGCATCTTGGAGAACTTAAATAAGACAGA	840
QY	889	TATCAAAATTCCTGAGGACTGGCCCTTACCAAGAAAGCTTCGACGCTTCTCAAGGAGCTAAT	948
- DB	841	TATCAAAATTCCTGAGGACTGGCCCTTACCAAGAAAGCTTCGACGCTTCTCAAGGAGCTAAT	900
QY	949	GTCACATGGATATTCCTGAGCTAAATGGACTGCACCTGTATGAGGAGGGTCTCATAGT	1008
DB	901	GTCACATGGATATTCCTGAGCTAAATGGACTGCACCTGTATGAGGAGGGTCTCATAGT	960
QY	1009	TTCTCTGTAAAGATAATGTTTCAACGAAGATCGGGTGACAAAGCCATAGAGAATC	1068
DB	961	TTCTCTGTAAAGATAATGTTTCAACGAAGATCGGGTGACAAAGCCATAGAGAATC	1020
QY	1069	AAATCTGCCAAGAAATAATCTGTCGAAGGAGACTCGAGTCCCTTTTCAAGCCAACTGCC	1128
DB	1021	AAATCTGCCAAGAAATAATCTGTCGAAGGAGACTCGAGTCCCTTTTCAAGCCAACTGCC	1080
QY	1129	ACCACATCAGCCGCTAAACGGAGGAGACTTCGGATATAACAAAGCAGGCGACTGCG	1188
DB	1081	ACCACATCAGCCGCTAAACGGAGGAGACTTCGGATATAACAAAGCAGGCGACTGCG	1140
QY	1189	AACAAGAAAAAAGGCTGGTGGAAAGAAAAATAATCTTGGATGCTTGATGTACAACTA	1248

Db 1141 AACAGAAACAAAGGCTGTGGAAAGAAATAATCTTGGATGCTTGTATACACTA 1200  
QY 1249 CGACTAGAAAGACGCGGTGGCTGATCCTTGCCTTAGATTAATTAACCTCCCTGTTTAA 1308  
Db 1201 CGACTAGAAAGACGCGGTGGCTGATCCTTGCCTTAGATTAATTAACCTCCCTGTTTAA 1260  
QY 1309 ACTCAGAGCTTTGGTAAAGCTTTCATGTTTCAAGCTGGGTAAAGTTAGTTGTGTTG 1368  
Db 1261 ACTCAGAGCTTTGGTAAAGCTTTCATGTTTCAAGCTGGGTAAAGTTAGTTGTGTTG 1320  
QY 1369 AAGAGATTGGTGTACCAAGTAACAAAACCTTATCGCTGTTTTT 1411  
Db 1321 AAGAGATTGGTGTACCAAGTAACAAAACCTTATCGCTGTTTTT 1363

## RESULT 5

US-08-455-968E-2  
; Sequence 2, Application US/08455968E  
; Patent No. 5874283  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, John L.  
; APPLICANT: Hsieh, Chih-Lin  
; APPLICANT: Lieber, Michael  
; TITLE OF INVENTION: Mammalian Flap-Specific Endonuclease  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/455, 968E  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Apple, Randolph T.  
; REGISTRATION NUMBER: 36,429  
; REFERENCE/DOCKET NUMBER: 18985-000100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-576-0200  
; TELEFAX: 415-576-0300  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1144 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
US-08-455-968E-2

Query Match 20.7%; Score 302.6; DB 2; Length 1144;  
Best Local Similarity 56.6%; Pred. No. 7.5e-70;  
Matches 601; Conservative 0; Mismatches 454; Indels 6; Gaps 2;

QY 85 ATGGGCGATCAAGGGTTTGACGAAACTGCTGGCGGACAAATCGCCCAAGCGCATGAAGGAG 144  
Db 1 ATGGGAATTCAGGCGCTGGCCAACTAATGCTGATGGCCCGCCAGTGCCATCCGGGAG 60  
QY 145 CAGAAGTTCAGAGCTACTTGGCGCCGCAAAATCCCGCTGCGACGACAGCATGATATAC 204  
Db 61 AATGACATCAAGAGCTACTTGGCGCGTAAAGGTGGCCCATTTGCTCTATGAGCATTTAT 120  
QY 205 CAGTTCCTGATTGATGTTGGNAGACAGGCATGGAACCTCTCACAATGAAGCTGGTCAA 264  
Db 121 CAGTTCCTGATTGTTGCTGCCAG---GGTGGGATGTGCTGCAGAAATGAGAGGGTGAG 177

QY 265 GTCACTAGTCATTTGCAAGGAATGTTCAACGGGCAATAAGATTACTTGGAGCGSGAATC 324  
Db 178 ACACACAGCCACCTGATGGGATGTTCTACCGCACCATTCCATGATGGAGAACGGCATC 237  
QY 325 AAGCCAGTTTATGTTTTTGTATGGCAAGCCTCCTGATATGAAGAAACAAGAGCTTGTCTAAA 384  
Db 238 AAGCCCGTGTATGTTTGTATGGCAAGCCTCCTGATATGAAGAAACAAGAGCTTGTCTAAA 297  
QY 385 AGATACATAAAGAGATGATGCAACCAAGATCTGACTGAGGCGAGTAGAGGTAGGAGAT 444  
Db 298 CGCAGTGAAGCGCGGCTGAGGCAGAGAAGCAGCTGCAGCAGGCTCAGGCTGTGGGGCC 357  
QY 445 AAGATCGGATTTGAAAAATTTGAGCAAGAGGACTCTAAAGGTCAACAAGGCAACAACGAA 504  
Db 358 GAGCAGAGGTGGAAAAATTTCACTAAGCGGCTGTGAAGGTCACTAAGCACCACATGAT 417  
QY 505 GATTGTAACGGCTATTAAAGACTTTATGGGGTTCCTGTTGTAGAGGACCTTCTTGAAGCA 564  
Db 418 GAGTCAAAACATCTGCTGAGCCTCATGGCATCCCTTATCTTTGATGCACCCAGTGAGGCA 477  
QY 565 GAAGCAGAATGTCAGCCCTTTGTCATAAAGATAAGGTGTCGCTGTTGCTTCAAGAGAT 624  
Db 478 GAGGCCAGCTGTGCTGCCCTGGTGAAGGTGGCAAGGTCTATGCTGGGCTACCGAGGAC 537  
QY 625 ATGACTCCCTTACTTTTGGGGCTCCACGGTTCCTTCTGTCATTTTAATGGATCCAAGTCC 684  
Db 538 ATGACTGCCCTACCTTCGGCAGCCCTGCTGCTAATGCGACACCTGACTGC-AGTGAAGCC 597  
QY 685 AAGAAATACCTGTGATGGAAATTTGATGTTGCCAAGGTTTGGAGGAGCTTGAACCTACC 744  
Db 598 AAAAGCTGCCAATCCAGGAATTCACCTGAGCGCGGATTCGTCAGGAGCTGGGCTGAAC 657  
QY 745 ATGACCAGTTCATTGATTTGTGATCTGCTGTGGATGTGACTATTTGATAGCATCAA 804  
Db 658 CAGGAACAGTTTGTGGATCTGTGATCTCTGCTAGGAGTGAATCTGCTGAGAGTATCCGG 717  
QY 805 GGATATCGGGGGCAACAGACTCTCAAACTTATTCGTCAACTGGGTCCATAGAAAGCATC 864  
Db 718 GGATATCGGGGGCAACAGACTCTCAAACTTATTCGTCAACTGGGTCCATAGAAAGCATC 777  
QY 865 TTGGAGAACTTATAAAGACAGATATCAAAATTCCTGAGGAGCTGGCTTACCAAGAGCT 924  
Db 778 GTGCGGGGAGTTGACCCCAACAAGTACCTGTGCGAGAAAATTTGGCTCCACAAGAGGCT 837  
QY 925 CGAGCTTTGTTCAAGGAGCCCTAATG---TCACATTGGATATTTCTGAGCTAAAATGGACT 981  
Db 838 CACCAGCTCTTCTTGGAACTGAGTGTGCTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 897  
QY 982 GCACCTGATGAGGAGGTCTCATAAAGTTTCTGTTAAAGATAATGTTTTCACGGAAGAT 1041  
Db 898 GAGCCAAATGAAGAAGAGCTGATCAAGTTCATGTGTGGTGAAGAGCAGTTCTCTGAGGAG 957  
QY 1042 CCGGTGACAAGGCCATAGAGAATCAAACTCTCCCAAGATAAATCGTCCCAAGGAAGA 1101  
Db 958 CGAATCCGAGTGGGGTCAAGAGGCTGAGTAAGAGCCGCCCAAGGACGACCCAGCGCGC 1017  
QY 1102 CTCGAGTCTTTTCAAGCCCAACTGCCACCATCATCAGCAC 1142  
Db 1018 CTGATGATTTCTTCAAGGTGACCGGCTCACTCTCTTCAGC 1058

## RESULT 6

US-08-455-968E-28  
; Sequence 28, Application US/08455968E  
; Patent No. 5874283  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, John L.  
; APPLICANT: Hsieh, Chih-Lin  
; APPLICANT: Lieber, Michael  
; TITLE OF INVENTION: Mammalian Flap-Specific Endonuclease  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,968E  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 18985-000100  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1144 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (polynucleotide)  
US-08-455-968E-28

Query Match 20.7%; Score 302.6; DB 2; Length 1144;  
Best Local Similarity 56.6%; Pred. No. 7.5e-70;  
Matches 601; Conservative 0; Mismatches 454; Indels 6; Gaps 2;

QY 85 ATGGGATCAAGGTTTGGAGAACTGCTGGCGACAATGCGCCCAAGCGGATGAAGAG 144  
DB 1 ATGGGAATTCAGGCCCTGGCCAACTAATTGCTGATGTGGCCCCCATGCGCATCCGGAG 60  
QY 145 CAGAAGTTCGAGAGCTACTTTCGGCCGAAAATCGCCGTCAGCCAGCATGAGCATATAC 204  
DB 61 AATGACATCAAGAGCTACTTTCGGCCGTAAGGTGCCCATGATGCTCTATGAGCATTTAT 120  
QY 205 CAGTTCCTGATTTAGTGAAGGACAGGATGGAAGTCTCACAATGAGTGGTGA 264  
DB 121 CAGTTCCTGATTTAGTGGTCCAG--GGTGGGATGCTGTCAGAAATGAGGGGTGAG 177  
QY 265 GTCACTAGTTCATTTGCAAGAAATTTCAACCGGACAATAAGATTACTTGAAGCGGAATC 324  
DB 178 ACCACGACCATGATGGGATGTTCTACCGCACCATTCGATGATGGAGACGGCATC 237  
QY 325 AAGCCAGTTTATGTTTGTGATGCAAGCCCTCCTGATATGAAGAACAGAGCTTGTCTAAA 384  
DB 238 AAGCCGCTGATGCTTTGATGCAAGCCGACAGCTCAAGTCAGCGAGCTGGCCAAA 297  
QY 385 AGATACATAAAGAGATGATGCAACCAAGATGCTACTGAGGAGTAGAGTAGAGAT 444  
DB 298 CGCAGTGAAGCGGGGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 357  
QY 445 AAGATCGGATGCAAAATTTAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 504  
DB 358 GAGCAGAGAGTGAAGAAATTTCACTAAGCGGCTGGTGAAGGTCACTAAGCAGCAGAT 417  
QY 505 GATTGTAAGCGGCTATTAAAGCTTTATGGGGGTTCTGTTGTAGAGGACCCCTTCTGAAGCA 564  
DB 418 GAGTGAACATCTGCTGAGCCCTCATGGCATCCCTTATCTTGTATGACCCAGTGAGGCA 477  
QY 565 GAAGCAGAAATGTCAGGCCCTTTGCATAAAGAGATGAGTGTGCTGTTTGCATCAGAGAT 624  
DB 478 GAGGCCAGCTGCTGCGCCCTGGTGAAGGCTGGCAAGGCTATGCTGCGGCTACCGAGGAC 537  
QY 625 ATGGACTCCCTTACTTTTGGGGTCCACGGTTCCTTCCTCATTTAATGATCAGAGTTC 684

DB 538 ATGACTGCCTCACCTTCGGCAGCCCTGTGCTAATGCGACACCTGACTGCCAGTGAAGCC 597  
QY 685 AAGAAATACCTGTGATGGAATTTGATGTTGCCAAGGTTTTGGAGGAGCTTTGAACCTAC 744  
DB 598 AAAAGCTGCCAATCCAGGAATTCACCTGAGCGGATTTCTGACGAGCTGGCTGAAC 657  
QY 745 ATGACAGCTTCATTTGATTTGTCATCTGCTGTGATGTCATTTGTCATGATGATGAT 804  
DB 658 CAGGAACAGTTTGTGATCTGTGATCTGCTAGGAGTGTGCTGCTGAGAGTATCCGG 717  
QY 805 GGTATCGGGGGCAACAGCTCTCAAACTTATTCGTCAACATGGTGCATAGAAAGCATC 864  
DB 718 GGTATTGGCCCAAGCGGCTGTGACCTCATCAGAAAGCACAAGAGCATCGAGAGATC 777  
QY 865 TTGGAGAAATCTTAATAAAGCAGATATCAAAATTCCTGAGGACTGGCTTACCAAGAAGCT 924  
DB 778 GTGCGGCACTTGACCCCAACAAGTACCTGTGCGCAGAAATTTGGCTCCACAGGAGCT 837  
QY 925 CGAGCTTTTCAAGGAGCCTAATG---TCACATTTGATATTCCTGAGCTTTAAATGAGCT 981  
DB 838 CACCAGCTCTTCTTGAACCTGAGGTGCTGGACCCAGAGTCTGTGGAGCTGAAGTGGAGC 897  
QY 982 GCACCTGATGAGGAGGCTCTCAAAAGTTTCTGCTGTAAGATATGTTTCAACGAAGAT 1041  
DB 898 GAGCAAAATGAAGAGAGCTGATCAAGTTTATGTTGTAAGAAAGCATCTTCTTGAGGAG 957  
QY 1042 CGGTGACAAAGGCTATGAGAGATCAAAATCTGCCAAGAAATAATCGTCCAAAGGAAGA 1101  
DB 958 CGAATCCGCACTGGGTCAAGAGGCTGAGTAAGAGCGCCCAAGGCGACGACCCAGGCGCG 1017  
QY 1102 CTCGAGTCTTTTCAAGCCAACTGCCACCATCATCAGAC 1142  
DB 1018 CTGGATGATTTCTTCAAGGTGACCGGCTCACTCTCTTTCAGC 1058  
RESULT 7  
US-08-455-968E-9  
; Sequence 9, Application US/08455968E  
; Patent No. 5874283  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, John L.  
; APPLICANT: Hsieh, Chih-Lin  
; APPLICANT: Lieber, Michael  
; TITLE OF INVENTION: Mammalian Flap-Specific Endonuclease  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/455,968E  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Apple, Randolph T.  
; REGISTRATION NUMBER: 36,429  
; REFERENCE/DOCKET NUMBER: 18985-000100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-576-0200  
; TELEFAX: 415-576-0300  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2033 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

5.

4:

QY 205 CAGTCTCTGATTTGCTGAGGACAGGATGAACTCTCACAATGAAGCTGGTGAA 264  
|||||  
Db 121 CAGTCTCTGATTTGCTGCTGAG--GGTGGGATGCTGCTGCAACAGGAGGGTGAG 177  
|||||  
QY 265 GTCACATGATGATTTGCAAGGATGTTCAACCGGACAAATAGATTACTGGAAGCGGGATC 324  
|||||  
Db 178 ACCACAGGCTGATGGGATGTTAT-----GGCAACCATCCGATGAGAATGGCAT 231  
|||||  
QY 325 AAGCCAGTTATGTTTATGATGGCAAGCTCTCTGATATGAAGAAACAAGACTTGTCTAAA 384  
|||||  
Db 232 AAGCTGTGTACTCTTTGATGCAACACCAACAGCTGAAGTACAGGCGACTGGCCAG 291  
|||||  
QY 385 AGATCTCAAAAAGAGATGATGCAACCAAGATCTGATGAGGCGATGAGAGTAGGAT 444  
|||||  
Db 292 CGCAGTGAGAGGCGCCGAGGCTGAGAGCAACTGCAGCAGGCTCAGGAGGCTGGGATG 351  
|||||  
QY 445 AAGATGCGATTTGAAAATTTGAGCAAGGAGCTGTAAGGTCACAAAGGCAACACAAAGAA 504  
|||||  
Db 352 GAGGAGGCTGAGAGTTTCCAAAGAGGCTCGTGAAGGTCACCAACCAACACAATGAT 411  
|||||  
QY 505 GATTGTAAAGCGCTATTAGACTTATGGGGTTCCTGTTGATAGGCGACCTTCAGAGCA 564  
|||||  
Db 412 GAGTGAACACACTCGTGAAGCTGATGGGCTACCTTACCTGTGATGACCCAGCGAGGCA 471  
|||||  
QY 565 GAAGCAGATGTCAGGCTTGTGCAATAAAGGATGAGGCTGCTGCTGTTGCTTCAAGAAGT 624  
|||||  
Db 472 GAGCCAGCTGTGCTGCCCTGCAAGGCTGCAAAAGCTATGCTGCGCGCACGAGGAC 531  
|||||  
QY 625 ATGACTCCCTTACTTTGGGGCTCCACGGTTCCTTCGTCATTTAATGATCCAGTTCC 684  
|||||  
Db 532 ATGACTGCCCTCACTTTTGGAGCCCGGCTGCTAATGCGACACTTAACGCGAGCTGAGGCC 591  
|||||  
QY 685 AAGAAATACCTGTGATGAAATTTGATGTTGCAAGGTTTGGAGGAGCTTGAACCTCAC 744  
|||||  
Db 592 AAGAAGCTGCCATCAAGAGTTCCATCTGAGCGCGTCTCGAGGAGCTGGTCTGAAC 651  
|||||  
QY 745 ATGACCACTTCAATTTGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 804  
|||||  
Db 652 CAGGAGCGTTTGTGATCTGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 711  
|||||  
QY 805 GGTATCGGGGGCAACAGCTCTGAACTTATTCGTCAACATGGGTGCTCAAGAGATC 864  
|||||  
Db 712 GGCATTTGGCGCCAGCGGCTGTGGATCTCATCCAGAAACATAAGAGATCGAGGAGATC 771  
|||||  
QY 865 TTGAGAACTCTTAATAAGACAGATATCAATTTCTGAGGAGCTGCGCTTACCAAGAAGCT 924  
|||||  
Db 772 GTGAGCGGCTGACCCGACCAAGTACCCGCTTCCAGAGAACTGCTCCCAAGGAAGCC 831  
|||||  
QY 925 CGAGCTTTGTTCAAGGAGCTTAA--TGTACATTTGGATATTCCTGAGCTAAAATGGACT 981  
|||||  
Db 832 CAGCAGCTCTTCTGAGGCGCAGAGTGTGAGCCAGAGCTCTGTGAGCTGAAAGTGGAGC 891  
|||||  
QY 982 GCACCTGATCAGGAGGCTCTCATAGCTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1041  
|||||  
Db 892 GAGCAATGAGAGAGTGTGTTCAATTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 951  
|||||  
QY 1042 CGGGTGCAAGGCGCATAGAGAGATCAATCTGCCAAGATAAATCTGCGCAAGGAAGA 1101  
|||||  
Db 952 CGAATTCGAGTGGGCTCAGCGGCTGAGTAAAGCGCGGAGCGGAGCAGCCAGGAGGCG 1011  
|||||  
QY 1102 CTCGAGTCTTTTCAAGCCCACTGCGACACATCAGCACCGCTAAAACGGAGGAGACT 1161  
|||||  
Db 1012 CTCGATGATTTCTAAGGTGACAGGCTCACCTCTCTCAGC-----TAAGGC 1059  
|||||  
QY 1162 TCGGATAAACAGCAGGCGCTGCGAACAAGAAACAAAGGCTGTTGGAAAGGAAGA 1221  
|||||  
Db 1060 AAGGAGCCAGAACCCCAAGGGGCTGCTAAGAGAAAGCAAGCAAGCTGGGGGAGG 1119  
|||||  
QY 1222 T 1222  
Db 1120 T 1120

RESULT 9  
US-08-455-968E-6  
; Sequence 6, Application US/08455968E  
; Patent No. 5874283  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, John L.  
; APPLICANT: Hsieh, Chih-Lin  
; APPLICANT: Lieber, Michael  
; TITLE OF INVENTION: Mammalian Flap-Specific Endonuclease  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/455,968E  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Apple, Randolph T.  
; REGISTRATION NUMBER: 36,429  
; REFERENCE/DOCKET NUMBER: 18985-000100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-576-0200  
; TELEFAX: 415-576-0300  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1149 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
US-08-455-968E-6  
Query Match 18.5%; Score 270.2; DB 2; Length 1149;  
Best Local Similarity 55.5%; Pred. No. 2.3e-61;  
Matches 599; Conservative 0; Mismatches 453; Indels 27; Gaps 3;  
QY 85 ATGGGCTCAAGGCTTACGAAACTGCTGGCGGACATGCGCCAAAGGCGATGAAGGAG 144  
|||||  
Db 1 ATGGGCTTAAAGGTTGAATGCAATTATATCGGAACATGTTCCCTCTGCTATCAGAAA 60  
|||||  
QY 145 CAGAACTTCGAGAGTACTTCGGCGGCAAAATCGCGTCGACGCGCAGCATGAGCATATAC 204  
|||||  
Db 61 AGCGATATCAAGAGCTTTTGGCAGAAAAGGTGGCATGCGCTCTATGCTCTATAT 120  
|||||  
QY 205 CAGTCTCTGATTTGATGTTGGAAGGACAGCATGGAACCTCTCACAATAAGCTGGTGAA 264  
|||||  
Db 121 CAGTTTTTAAATTCGTGTAAGACAGCAACCGTGGCGAGTTGACCATGAAGCGGTGAA 180  
|||||  
QY 265 GTCACATGATTTGCAAGGAATGTTCAACCGGACAAATAGATTACTGGAAGCGGGATC 324  
|||||  
Db 181 ACAACGCTACACTTGTGTTGTTTATAGGACACTGAGAATGATTGATTAACGGTATC 240  
|||||  
QY 325 AAGCCAGTTTATGTTTGTGATGGCAAGCTCTCTGATATGAAGAAACAAGAGCTTGTCTAAA 384  
|||||  
Db 241 AAGCCTTTGATGTTCTGCGCGGCAAACTCCAGCTTTGAATCTCATGATTGACAAAG 300  
|||||  
QY 385 AGATCTCAAAAAGAGATGATGCAACCAAGATCTGACTGAGGCGAGTAGAGTAGAGAT 444  
|||||  
Db 301 CGGTCTTCAAGAGGCTGGAAACAGAAAAAACTGGCAGAGGCAACACAGAAATTGGAA 360  
|||||  
QY 445 AAAGATGCGATTGAAAAATTGAGCAAGAGGACTGTAAAGGTCAACAAGGCAACACAGGAA 504  
|||||  
Db 361 AAGATGAACGAAGAAGA-----AGATTGTTGAAGGCTCAAAAAGACATAATGAA 411  
|||||

QY 505 GATTGTAACGGCTATTAGAGCTTATGGGGTTCCTGTTCTAGAGCAGCCTCTTCTGAAGCA 564  
Db 412 GAAGCCCAAAATTAAGAGCTTATGGGATCCCATATATATAGCGCAAGGAGCT 471  
QY 565 GAAGCAGATGTGACAGCCCTTTTCATAAACGATTAAGGTGTTCGCTGTGCTTCAGAAAT 624  
Db 472 GAGGCTCAATGTCTGAGTTGGCAAGAAAGGAAAGGTGTATGCCGAGCAAGTGAAGAT 531  
QY 625 ATGAGCTCCCTTACTTTTGGGGTCCACAGGTTCCCTGCTGCTCATTTATGGATCAAGTTC 684  
Db 532 ATGACACACACTGTATAGAACACCTTCTGTTGAGACATTTGACTTTTCAGAGGCC 591  
QY 685 AAGAAATACCTCTGATGGAATTTGATGTTGCCAAGTTCCTGAGAGCTTGAACATCACC 744  
Db 592 AAGAGGACCGATTCACGAAATAGATGATGAAATAGTTTGTAGAGACTCGACTTGACA 651  
QY 745 ATGACAGCTTCATTTGATGTCATCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 804  
Db 652 ATAGAGCAGTTTGTGATCTTTCATTAATGCTTGGTGTGACTACTGTGAAGCATCAGA 711  
QY 805 GGTATCGGGGGCAACAGCTCTGAACCTTATTCGTCACATGGGTCATAGAAAGCATC 864  
Db 712 GGTGTTGGTCCAGTGACAGCCTTAAATTTGATAAAACGATGATCCATCGAAAAATC 771  
QY 865 TTGGAGAATCTT-----ATAAAGACAGATATCAAAATTCCTGAGGACTGG 909  
Db 772 GTGAGTTTATTAATCTGGGAGTCAAAACACTAAATGGAAATCCAGAGACTGG 831  
QY 910 CTTTACCAAGAGCTCGACCTTTGTTCAAGGAGCCTTAATGTCATTTGGATA---TTCCT 966  
Db 832 CTTTACCAAGAGCTCGACCTTTGTTCAAGGAGCCTTAATGTCATTTGGATA---TTCCT 966  
QY 967 GAGCTAAATGAGCTGACCTGTATGAGGAGGCTCTCATAGTTTCTGTTAAAGATAAT 1026  
Db 892 AACTTGAATGGTGGCCACCAAGAGAGAACTTATCGAGTATTTATGATGATAAG 951  
QY 1027 GGTTCACCAAGAGCTGGTGACAAAGGCTATAGAGAGATCAAAATCTGCCAAGATAAA 1086  
Db 952 AATTCAGTGAAGAGAGTAAATCTGGTATATCAAGTTGAAAGAGCTTGAATCT 1011  
QY 1087 TCGTCGAAGAGAGCTGAGTCTCTTTTCAAGCCCACTGCCACCATCAGCACCGCT 1145  
Db 1012 GGCATTGAGGTTAGATGGTCTTCCAAAGTGTGCTTAAGACAAAGGACAGCT 1070

## RESULT 10

US-08-757-653-175  
; Sequence 175, Application US/08757653  
; Patent No. 5843669

## GENERAL INFORMATION:

; APPLICANT: Kaiser, Michael W.  
; APPLICANT: Lyamichev, Victor I.  
; APPLICANT: Lyamichev, Natasha  
; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using  
; TITLE OF INVENTION: Thermostable FEN-1 Endonucleases  
; NUMBER OF SEQUENCES: 190  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Medien & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/757,653  
; FILING DATE:  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: FORS-02565  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 175:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1023 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1020  
; US-08-757-653-175

Query Match 10.9%; Score 159.6; DB 2; Length 1023;  
Best Local Similarity 54.6%; Pred. No. 2e-32;  
Matches 318; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 134 CGATGAAGGAGCAGAAAGTTTCGAGAGCTTCTCGGCCCCAAATTCGCGTCCACGCCAGCA 193  
Db 29 CAAGAAAAAGAAATAGCTTAGAAAAACCTATACGGGAAAAAATCGCAATCGACCTCTTA 88  
QY 194 TGACATATACCAAGTTCCTGATTTGAGTGTGAGGAGCAGGATGGAACCTTCACAAATG 253  
Db 89 ATCAATCTACCAATTTTGTCCACAATAAGACAGAAAGATGGAACTCCCTTAAGGATT 148  
QY 254 AAGCTGTGAGTCACTAGTCAATTTGCAAGGAATGTTCAACCGGACAAATAGATTACTGG 313  
Db 149 CAAGGCTAGATAACCTCCACCTAAGCGGGCTCTTTTACAGGACAAATAACCTATGG 208  
QY 314 AAGCGGGAATCAAGCCAGTTTATGTTTGTATGCGCAAGCCCTCTCTGATATGAAGAAACAAG 373  
Db 209 AGGCTGAATAAAACCTGTGTATGTTTGTATGAGAAACCTCCAGAAATTCAAAAAGAAAG 268  
QY 374 AGCTTGTAAAAGTACTCAAAAAGAGATGATGCAACCAAGATCTGACTGAGGAGTAG 433  
Db 269 AGCTGAAAAAAG 328  
QY 434 AGGTAGGAGATAAAGATGCGATTGCAAAATTTAGCAAGAGAGACTGTAAAGTCAACAAGC 493  
Db 329 AAAAGAGAGATAG 388  
QY 494 AACACAACGAAGATTGTAACGGCTATTAAAGACTTTATGGGGTTTCTGTTCTAGAGGAC 553  
Db 389 TGCTCATCGAGGATGCAAAAAACTCTTAGAGCTTATGGGAATTCCTATATAGTTCACAGC 448  
QY 554 CTCTGTAAGCAGAGCAGAAATGTGACGCCCTTTTCATATAAAGATAGGTTCTCGCTGTTG 613  
Db 449 CTAGCAGGAG 508  
QY 614 CTTCAGAAAGATATGAGACTCCCTTACTTTTGGGGTTCACAGGTTTCTCTGCTCATTAATGG 673  
Db 509 CTAGTCAAGATTAGGATTCCTACTTTTGGAGCTCCAGAGCTTGTAGAACTTAACAA 568  
QY 674 ATCCAAAGTTCAGAAAAATACCTGTGATGGAATTTGATGTTG 715  
Db 569 TAACAGGAAAAAGAAAGTTGCTCGGGAAAAATGTCTACGTG 610

## RESULT 11

US-08-823-516-78  
; Sequence 78, Application US/08823516  
; Patent No. 5994069  
; GENERAL INFORMATION:  
; APPLICANT: Hall, Jeff G.  
; APPLICANT: Lyamichev, Victor I.  
; APPLICANT: Mast, Andrea L.  
; APPLICANT: Brow, Mary Ann D.

```

; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
; TITLE OF INVENTION: Sequential Invasive Cleavages
; NUMBER OF SEQUENCES: 163
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,516
; FILING DATE: 24-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01072
; FILING DATE: 21-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/759,038
; FILING DATE: 02-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/758,314
; FILING DATE: 02-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/756,386
; FILING DATE: 29-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02736
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 78:
; LENGTH: 1023 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1020
; US-08-823-516-78

```

```

Query Match      10.9%; Score 159.6; DB 2; Length 1023;
Best Local Similarity 54.6%; Pred. No. 2e-32;
Matches 318; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 134 CGATGAAGGAGCAGACGCTCGAGAGCTACTCGGCCGCAAAATCGCGTCGACGCCAGCA 193
Db 29 CAGAAAAGAAATGAGTTAGAAAACCTATACGGGAAAAAATCGCAATCGAGCTCTTA 88
QY 194 TGACATATACCAAGTTCCTGATTGTAGTTGGAAGGACGAGCAGTGAACCTCTCACAAATG 253
Db 89 ATGCAATCTACCAATTTTGTCCACATAAGACAGAAAGATGGAACCTCCACTTATGATT 148
QY 254 AAGCTGTGAAGTCACTAGTCTATTTCGAAGGAATGTTCAACCGGCAATTAAGATTCTGG 313
Db 149 CAAAGGAGTAGAATAACCTCCACCTAAGCGGCTCTTTTACAGGACAAATAAACCTAATGG 208
QY 314 AAGCGGAATCAAGCCAGTTTATGTTTTCATGGCAAGCCTCTCTGATATGAAGAACAG 373

```

```

Db 209 AGGTGGAATAAAACCTGTGTATGTTTTTGTGAGAACCTCCAGAAATTCAAAAGAAAG 268
QY 374 AGCTTGCTAAAGATACTCAAAAAGAGATGATGCAACCAAGATCTGACTCAGGCAGTAG 433
Db 269 AGCTCGAAAAGAGAGAGAGAGAGAGAGAGCTGAAGAAAAGTGGAGAGAACCACTTG 328
QY 434 AGGTAGGAGATAAAGATCCGATTGAAAATTTGACGAAGAGGACTGTAAAGTCAACAGGC 493
Db 329 AAAAGGAGAGATAGAGGAGCAAGCAAAATATGCCAAAGAGCAACCGGTTAATGAAA 388
QY 494 AACACAGGAAGATTGTAACGGCTATTAAAGCTATTAGACATTATGGGGGTTCTCTGTTTAGAGGCAC 553
Db 389 TGCTCATCGAGGATGCAAAAACCTCTTAGAGCTTATGGGAATTCCTATAGTTCAGGCAC 448
QY 554 CTCTGGAAGCAGAGCAGAAATGTCAGCCCTTTTCATATAACGATAAGTGTTCCTGTTG 613
Db 449 CTACGAGGAGAGAGCCCAAGCTGCATATATGGCCGCAAGAGGAGCGGTGATGCAATCGG 508
QY 614 CTTCAGAAGATATGGACTCCCTTACTTTTGGGGTCCACCGTTCTCTTCGTCATTAAATGG 673
Db 509 CTAGTCAAGATTACGATTCCCTACTTTTGGAGCTCCAAGACTTTGTAGAAACTTAACAA 568
QY 674 ATCCAAGTTCACAGAAAATACCTGTGATGGAATTTGATGTTG 715
Db 569 TAACAGGAAAAGAAAGTTGCTCTGGGAAAATGTTCTACGTCG 610

```

```

RESULT 12
US-08-759-038-114
; Sequence 114, Application US/08759038
; Patent No. 6090543
; GENERAL INFORMATION:
; APPLICANT: Prudent, James R.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; APPLICANT: Dahlberg, James E.
; TITLE OF INVENTION: Cleavage Of Nucleic Acids
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,038
; FILING DATE: 02-DEC-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 29-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 114:

```



## SEQUENCE CHARACTERISTICS:

LENGTH: 1023 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1020  
US-08-759-038-114

Query Match 10.9%; Score 159.6; DB 3; Length 1023;

Best Local Similarity 54.6%; Pred. No. 2e-32;

Matches 318; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 134 CGATGAAGGAGCAGAGTTTCGAGAGCTACTTCGCCCGCAAAATCGCGTCGACGCCAGCA 193  
Db 29 CAAGAAAGAAATTTGATTAGAAAACCTATACGGGAAAAAAATCGCAATCGACGCTCTTA 88  
QY 194 TGAGCATATACCAAGTTCCTGATTGTAGTTGGAAGCAGGCATGGAACCTCTCAAAATG 253  
Db 89 ATGCATCTACCAATTTTTCACATAAAGACAGAAAGATGGAATCCACTTATGGATT 148  
QY 254 AGCTGGTGAAGTACATTTTCGACGAAATGTTCAACCGGACAAATTAAGATTACTGG 313  
Db 149 CAAAGGTAGATAAATCCACCTAAGCGGCTCTTTTACAGGACAAATAAACCTAATGG 208  
QY 314 AACGGGAATCAAGCAGTTTATGTTTTCGACAGCCCTCTGATATGAAGAAACAG 373  
Db 209 AGCTGGAATAAACCCTGTATGTTTTTATGAGAACCTCCAGAAATTCAAAAGAAAG 268  
QY 374 AGCTTGCTAAAGATCTCAAAAGAGATGATGCAACCAAGATCTGACTGAGCGAGTAG 433  
Db 269 AGCTGGAATAAACCCTGTATGTTTTTATGAGAACCTCCAGAAATTCAAAAGAAAG 328  
QY 434 AGTGAAGATAAAGATGTAAGCGCTATTAAAGCTTATGGGGTTCCTGTTCTAGAGCAC 493  
Db 329 AAAAAGGAGAGATAGAGGAAGCAAGAAATATGCCAAAGACCAACCGGTAATGAAA 388  
QY 494 AACACAAAGAGATGTAAGCGCTATTAAAGCTTATGGGGTTCCTGTTCTAGAGCAC 553  
Db 389 TGCATCGAGGATGCAAAACCTCTTATGAGCTTATGGGAATTCCTATGATCAATGG 673  
QY 554 CTTCGAGGAGAGCCCAAGCTGCATATATGCGCCGCAAGAGGAGCGTGTATGCAATGG 613  
Db 449 CTAGCGAGGAGAGCCCAAGCTGCATATATGCGCCGCAAGAGGAGCGTGTATGCAATGG 508  
QY 614 CTTCGAGGATGAGTCCCTTACTTTTGGGGCTCCACGGTTCCTTCGTCATTTAATGG 673  
Db 509 CTAGTCAAGATTAAGATTCCTACTTTTGGAGCTCCAAAGACTTGTAGAACTTAACAA 568  
QY 674 ATCCAAAGTTCAGAAATACCTGTGATGGAATTTGATGTTG 715  
Db 569 TACAGGAAAAGAAAGTTCCTCGGGAAAAATGCTACGTCG 610

## RESULT 13

US-08-758-314-114

Sequence 114, Application US/08758314

Patent No. 6090606

## GENERAL INFORMATION:

APPLICANT: Kaiser, Michael W.  
APPLICANT: Lyamichev, Victor I.  
APPLICANT: Lyamichev, Natasha  
TITLE OF INVENTION: Improved Cleavage Agents  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:

ADDRESS: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/758,314  
FILING DATE: 02-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/  
FILING DATE: 29-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/682,853  
FILING DATE: 12-JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,491  
FILING DATE: 24-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02575  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 114:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1023 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1020  
US-08-758-314-114

Query Match 10.9%; Score 159.6; DB 3; Length 1023;

Best Local Similarity 54.6%; Pred. No. 2e-32;

Matches 318; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 134 CGATGAAGGAGCAGAGTTTCGAGAGCTACTTCGCCCGCAAAATCGCGTCGACGCCAGCA 193  
Db 29 CAAGAAAGAAATTTGATTAGAAAACCTATACGGGAAAAAAATCGCAATCGACGCTCTTA 88  
QY 194 TGAGCATATACCAAGTTCCTGATTGTAGTTGGAAGCAGGCATGGAACCTCTCAAAATG 253  
Db 89 ATGCATCTACCAATTTTTCACATAAAGACAGAAAGATGGAATCCACTTATGGATT 148  
QY 254 AAGCTGGTGAAGTTCATGCTAGTTCATTTTCAAGGAATGTTCAACCGGACAAATTAAGATTACTGG 313  
Db 149 CAAAGGTAGATAAATCCACCTAAGCGGCTCTTTTACAGGACAAATAAACCTAATGG 208  
QY 314 AAGCGGGAATCAAGCAGTTTATGTTTTCGACAGCCCTCTGATATGAAGAAACAG 373  
Db 209 AGGCTGGAATAAACCCTGTATGTTTTTATGAGAACCTCCAGAAATTCAAAAGAAAG 268  
QY 374 AGCTTGCTAAAGATCTCAAAAGAGATGATGCAACCAAGATCTGACTGAGCGAGTAG 433  
Db 269 AGCTGGAATAAAGAGAGGAGGAGAGAGCTGGAAGAAAGTGGAGAGAGCACTTG 328  
QY 434 AGGTAGGAGATAAAGATGCGATTGAAAAATTTAGCAAGAGAGCTGTAAAGGTCTACAAGGC 493  
Db 329 AAAAAGGAGAGATAGAGGAGCAAGAAATATGCCCAAGAGCAACCGGTAATGAAA 388  
QY 494 AACACAAAGAGATGTAAGCGCTATTAAAGCTTATGGGGTTCCTGTTCTAGAGCAC 553  
Db 389 TGCATCGAGGATGCAAAACCTCTTATGAGCTTATGGGAATTCCTATGATCAAGCAC 448  
QY 554 CTTCGAGCAGACAGAGATGTCAGCCCTTTGACATAAAGCTTTCGCTGCTTGGCTTGG 613  
Db 449 CTAGCGAGGAGAGCCCAAGCTGCATATATGCGCCGCAAGAGGAGCGTGTATGCAATGG 508

Qy	614	CTTCAGAGATATGGACTCCCTTACTTTTGGGCGCTCCACGGTTCCTTGGTCATTTAAATGG	673
Db	509	CTAGTCAAGATTACGATTCCCTACTTTTTTGGAGCTCCAGACTTGTATGAACCTTAACAA	568
Qy	674	ATCCAAGTTCACAGAAAATACCTGTGTATGGGAATTTGATGTG	715
Db	569	TAACAGAGAAAAGAAAGTGTGCCCTGGGAAAATGTCTACGTG	610

## RESULT 14

```

US-09-146-319-1
? Sequence 1, Application US/09146319
? Patent No. 6251649
? GENERAL INFORMATION:
? APPLICANT: Matsui, Ikuo
? APPLICANT: Ishikawa, Kazuhiko
? APPLICANT: Kosugi, Yoshitsugu
? APPLICANT: Matsui, Eriko
? APPLICANT: Kawasaki, Satoko
? TITLE OF INVENTION: THERMOSTABLE FLAP ENDONUCLEASE DERIVED FROM A
? TITLE OF INVENTION: HYPERTHERMOPHILE BACTERIUM BELONGING TO THE GENUS
? TITLE OF INVENTION: PYROCOCCLUS
? FILE REFERENCE: 081356-0121
? CURRENT APPLICATION NUMBER: US/09/146,319
? CURRENT FILING DATE: 1998-09-03
? EARLIER APPLICATION NUMBER: JP 239440/1997
? EARLIER FILING DATE: 1997-09-04
? NUMBER OF SEQ ID NOS: 7
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 1
? LENGTH: 1032
? TYPE: DNA
? ORGANISM: Pyrococcus horikoshii
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1)..(1029)

```

[illegible]

RESULT 15

```

US-09-175-973-1
; Sequence 1, Application US/09175973A
; Patent No. 6255081
; GENERAL INFORMATION:
; APPLICANT: MATSUI, Ikuo
; APPLICANT: ISHIKAWA, Kazuhiko
; APPLICANT: KOSUGI, Yoshitsugu
; APPLICANT: MATSUI, Eriko
; APPLICANT: KAWASAKI, Satoko
; TITLE OF INVENTION: THERMOSTABLE FLAP ENDONUCLEASE DERIVED FROM A
; TITLE OF INVENTION: HYPERTHERMOPHILE BACTERIUM BELONGING TO THE GENUS
; TITLE OF INVENTION: PYROCOCOCCUS
; FILE REFERENCE: 081356/0126
; CURRENT APPLICATION NUMBER: US/09/175, 973A
; CURRENT FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Pyrococcus horikoshii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1029)

```

QY 614 CTTCAAGATATGGACTCCCTTACTTTTGGGGCTCCACGGTTCCTCGTCATTTAATGG 673  
Db 509 CGAGTCAAGATTATGATTCACCTACTCTTTGGTGCTCCCAAGTTGATTAGGATCTGACAA 568  
QY 674 ATCCAAGTTCCAAGAAAATACCTGTGTGATGGAATTTGATGTTG 715  
Db 569 TTACGGGAAAAGAAAGATGCTGGGAAGATGTTTACGTTG 610

Search completed: May 8, 2003, 21:36:06  
Job time : 101 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 19:52:22 ; Search time 195 Seconds  
(without alignments)  
9323.473 Million cell updates/sec

Title: US-09-805-311-1

Perfect score: 1463

Sequence: 1 caccgagaatagctcgccgccc.....aaaaaaaaaaaaaaaaaaaaa 1463

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1463	100.0	1463	10 US-09-805-311-1	Sequence 1, Appli
2	1421.8	97.2	1478	10 US-09-805-311-7	Sequence 7, Appli
3	1409.8	96.4	1541	10 US-09-805-311-3	Sequence 3, Appli
4	1350.2	92.3	1381	10 US-09-805-311-5	Sequence 5, Appli
5	159.6	10.9	1023	9 US-10-033-297-78	Sequence 78, Appli
6	159.6	10.9	1023	9 US-09-940-244-78	Sequence 78, Appli
7	152.8	10.4	1054	9 US-09-940-244-281	Sequence 281, Appli
8	151.4	10.3	485	9 US-09-918-995-30803	Sequence 30803, A
9	149.6	10.2	1115	9 US-09-940-244-289	Sequence 289, Appli
10	139.6	9.5	510	10 US-09-864-761-16172	Sequence 16172, A
11	135.6	9.3	1032	9 US-09-940-244-356	Sequence 356, Appli
12	133	9.1	1056	9 US-09-940-244-360	Sequence 360, Appli
13	128	8.7	514	9 US-09-940-244-287	Sequence 287, Appli
14	127.6	8.7	1023	9 US-09-940-244-383	Sequence 383, Appli
15	126.4	8.6	1056	9 US-09-940-244-336	Sequence 336, Appli
16	126.4	8.6	1164	9 US-09-940-244-272	Sequence 272, Appli
17	125.6	8.6	1115	9 US-09-940-244-275	Sequence 275, Appli
18	124	8.5	1041	9 US-09-940-244-378	Sequence 378, Appli
19	121	8.3	1023	9 US-09-940-244-401	Sequence 401, Appli

20	112.6	7.7	1056	9 US-09-940-244-340	Sequence 340, Appli
21	110.2	7.5	889	9 US-09-940-244-271	Sequence 271, Appli
22	108.2	7.4	1023	9 US-09-940-244-364	Sequence 364, Appli
23	108	7.4	1053	9 US-09-940-244-344	Sequence 344, Appli
24	106.2	7.3	981	9 US-09-940-244-352	Sequence 352, Appli
25	98.6	6.7	1011	9 US-09-940-244-164	Sequence 164, Appli
26	97.8	6.7	1008	9 US-09-940-244-348	Sequence 348, Appli
27	97.6	6.7	1164	9 US-09-940-244-286	Sequence 286, Appli
28	96	6.6	1062	9 US-09-940-244-374	Sequence 374, Appli
29	96	6.5	1071	9 US-09-940-244-388	Sequence 388, Appli
30	95.2	6.5	1017	9 US-09-940-244-369	Sequence 369, Appli
31	95.2	6.5	1017	10 US-09-777-430A-47	Sequence 47, Appli
32	94.8	6.5	386	9 US-09-940-244-276	Sequence 276, Appli
c 33	90	6.2	546	10 US-09-864-761-13411	Sequence 13411, A
34	89.2	6.1	981	9 US-10-033-297-74	Sequence 74, Appli
35	89.2	6.1	981	9 US-09-940-244-74	Sequence 74, Appli
c 36	86.6	5.9	245	10 US-09-864-761-32666	Sequence 32666, A
37	85.6	5.9	1053	9 US-09-940-244-393	Sequence 393, Appli
38	76.2	5.2	777	9 US-09-940-244-260	Sequence 260, Appli
39	74.6	5.1	840	9 US-09-940-244-274	Sequence 274, Appli
40	69.4	4.7	987	9 US-09-940-244-264	Sequence 264, Appli
41	67.8	4.6	296	9 US-09-940-244-273	Sequence 273, Appli
42	67.2	4.6	985	9 US-09-940-244-397	Sequence 397, Appli
43	66	4.5	514	9 US-09-940-244-282	Sequence 282, Appli
44	58	4.0	668	9 US-09-940-244-278	Sequence 278, Appli
c 45	51.6	3.5	418	10 US-09-960-352-4845	Sequence 4845, Ap

#### ALIGNMENTS

RESULT 1

US-09-805-311-1

; Sequence 1, Application US/09805311

; Patent No. US20010034886A1

; GENERAL INFORMATION:

; APPLICANT: Mahajan, Pramod B.

; TITLE OF INVENTION: Rad2/FEN-1 Orthologues and Uses

; FILE REFERENCE: 0961D

; CURRENT APPLICATION NUMBER: US/09/805,311

; CURRENT FILING DATE: 2001-03-13

; PRIOR APPLICATION NUMBER: 09/426,557

; PRIOR FILING DATE: 1999-10-22

; PRIOR APPLICATION NUMBER: 60/112,332

; PRIOR FILING DATE: 1998-12-15

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 1463

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (85)...(1221)

US-09-805-311-1

Query Match 100.0%; Score 1463; DB 10; Length 1463;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACGAGAATAGCTCGCGCGCGGTTTCTTGGCCACTCCGGCTCAGCGCGCGCGCGCA 60

|||||

Db 1 CACGAGAATAGCTCGCGCGCGGTTTCTTGGCCACTCCGGCTCAGCGCGCGCGCGCA 60

|||||

Qy 61 CCCGCCACAGCGCGCGAGAGATGGGATCAAGGGTTTGAGAACTCTCTGGCGGAC 120

|||||

Db 61 CCCGCCACAGCGCGCGAGAGATGGGATCAAGGGTTTGAGAACTCTCTGGCGGAC 120

|||||

Qy 121 AATCGGCCAAGCGATGAAGSACAGAGTTCAGAGCTACTTCGCGCGCAATCGCC 180

|||||

Db 121 AATCGGCCAAGCGATGAAGSACAGAGTTCAGAGCTACTTCGCGCGCAATCGCC 180

Qy	181	GTGACGCCAGCATGAGCATATACCAAGTTCCCTGATGTGATGTTGGAAGGACAGGCAATGGAA	240
Db	181	GTGACGCCAGCATGAGCATATACCAAGTTCCCTGATGTGATGTTGGAAGGACAGGCAATGGAA	240
Qy	241	ACTCTCAAAATCAAGCTGGTGAAGTCACTAGTCTATTTGCAAGGAATGTTCAACCCGGACA	300
Db	241	ACTCTCAAAATCAAGCTGGTGAAGTCACTAGTCTATTTGCAAGGAATGTTCAACCCGGACA	300
Qy	301	ATAAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTGTGATGGCAAGCCTCTGTAT	360
Db	301	ATAAGATTCTGGAAGCGGGAATCAAGCCAGTTTATGTTTGTGATGGCAAGCCTCTGTAT	360
Qy	361	ATGAAGAAACAAGAGCTTGCTTAAAGATACTCAAAAAGAGATGATGCAACCAAAAGATCTG	420
Db	361	ATGAAGAAACAAGAGCTTGCTTAAAGATACTCAAAAAGAGATGATGCAACCAAAAGATCTG	420
Qy	421	ACTGAGGCAGTAGAGTAGAGATAAAGATGGGATTTGAAAATTTGAGCAAGAGACTGTGA	480
Db	421	ACTGAGGCAGTAGAGTAGAGATAAAGATGGGATTTGAAAATTTGAGCAAGAGACTGTGA	480
Qy	481	AAGGTCAAAAGGCAACAACAAGAAATTTGTAACGGCTATTAAAGACTTTATGGGGTTCCCT	540
Db	481	AAGGTCAAAAGGCAACAACAAGAAATTTGTAACGGCTATTAAAGACTTTATGGGGTTCCCT	540
Qy	541	GTGTGAGAGCACCTTCTGAAGCAGAGCAGAAATGTGCAGGCCCTTTCGATTAAGCATAAG	600
Db	541	GTGTGAGAGCACCTTCTGAAGCAGAGCAGAAATGTGCAGGCCCTTTCGATTAAGCATAAG	600
Qy	601	GTGTTGCTGCTTGCTTCTCAGAAGATATGGACTCCCTTACTTTTGGGGCTCCACGGTTCCCT	660
Db	601	GTGTTGCTGCTTGCTTCTCAGAAGATATGGACTCCCTTACTTTTGGGGCTCCACGGTTCCCT	660
Qy	661	CGTCATTTAATGATCCAGTTCCAGAAATATACCTGTGATGGATTTGATTTGTTGCCAAG	720
Db	661	CGTCATTTAATGATCCAGTTCCAGAAATATACCTGTGATGGATTTGATTTGTTGCCAAG	720
Qy	721	GTTTGTGAGGAGCTTGAACTCACTGACCTGACCTGATTTGATTTGTGCATCCTGTGTGA	780
Db	721	GTTTGTGAGGAGCTTGAACTCACTGACCTGACCTGATTTGATTTGTGCATCCTGTGTGA	780
Qy	781	TGTGACTATTGTGATAGCATCAAAAGTATCGGGGGCAACAGACTCTGAAACTTATTTCGT	840
Db	781	TGTGACTATTGTGATAGCATCAAAAGTATCGGGGGCAACAGACTCTGAAACTTATTTCGT	840
Qy	841	CAACATGGGTCCATAGAAAGCATCTTTGGAGAATCTTAATAAGACAGATATCAAAATTCCT	900
Db	841	CAACATGGGTCCATAGAAAGCATCTTTGGAGAATCTTAATAAGACAGATATCAAAATTCCT	900
Qy	901	GAGGACTGGCCCTTACCAAGAAGCTGCACGCTTGTTCAAGGAGCCCTAATGTCACATTTGGAT	960
Db	901	GAGGACTGGCCCTTACCAAGAAGCTGCACGCTTGTTCAAGGAGCCCTAATGTCACATTTGGAT	960
Qy	961	ATTCTTGAGCTTAAATTTGGACTGCACCTGATGAGGAGGTCTCATAAAGTTTCTGGTAAAA	1020
Db	961	ATTCTTGAGCTTAAATTTGGACTGCACCTGATGAGGAGGTCTCATAAAGTTTCTGGTAAAA	1020
Qy	1021	GATAATGTTTCAACGAAGATCGGTTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAG	1080
Db	1021	GATAATGTTTCAACGAAGATCGGTTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAG	1080
Qy	1081	ATAAATCTGTGCAAGAGACTTCGAGTCCCTTTTTCAGGCCAAGTGCACACATCATGACA	1140
Db	1081	ATAAATCTGTGCAAGAGACTTCGAGTCCCTTTTTCAGGCCAAGTGCACACATCATGACA	1140
Qy	1141	CCGTAAACCGGAAGGAGACTTCGGATATAAACCAAGCAAGGAGCTGCCAACAGAAAAACA	1200
Db	1141	CCGTAAACCGGAAGGAGACTTCGGATATAAACCAAGCAAGGAGCTGCCAACAGAAAAACA	1200
Qy	1201	AAGCGTGTGGAAGAAGAAATAATCTTGGATGCTTGTATGTACAACTACGACTACGAAAG	1260
Db	1201	AAGCGTGTGGAAGAAGAAATAATCTTGGATGCTTGTATGTACAACTACGACTACGAAAG	1260
Qy	1261	CAGCGTGGCGTGATCACCTCGCTTAGATTAATTAACCTCCCTGTTTAACTACACGAGCTTT	1320

## RESULT 2

```

US-09-805-311-7
; Sequence 7, Application US/09805311
; Patent No. US20010034886A1
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Rad2/FEN-1 Orthologue
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0961D
; CURRENT APPLICATION NUMBER: US/09/805,311
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 09/426,557
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/112,332
; PRIOR FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1478
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (97)...(1233)
US-09-805-311-7

```

Query Match 97.28; Score 1421.8; DB 10; Length 1478;

Qy	7	AATAGCTCGCGCGCGGGTTCTTTGGCCACTCCGGCTACGCGCGCGCGCCACCGCC	66
Db	19	AATAGCTCGGGTCGCGGTTTCTTTGGCCACTCCGGCTACGCGCGCGCGCCACCGCC	78
Qy	67	ACACCCCGCAGACGAGATGGGCATCAAGGGTTTGACGNAACTGCTGGCGCAATGCG	126
Db	79	ACACCCCGCAGACGAGATGGGCATCAAGGGTTTGACGNAACTGCTGGCGCAATGCG	138
Qy	127	CCCAAGCGGATGAAGGAGCAGAAGTTCCGAGAGCTACTTTCGGCGCGCAAAATCGCGCTCGAC	186
Db	139	CCCAAGCGGATGAAGGAGCAGAAGTTCCGAGAGCTACTTTCGGCGCGCAAAATCGCGCTCGAC	198
Qy	187	GCCAGCATGACGATATACCAGTTCCTGATTGTAGTTGGAAAGGACAGGCATCGAATCTTC	246
Db	199	GCCAGCATGACGATCTACCAGTTCTCTGATAGTAGTTGGAAAGGACAGGCATGGAAATCTC	258
Qy	247	ACAAATGAGCTGGTGAAGTCACCTAGTCATTTCCAGAGGAATGTTCAACCGGACAAATAAG	306
Db	259	ACAAATGAGCTGGTGAAGTCACCTAGTCATTTCCAGAGGAATGTTCAACCGGACAAATAAG	318
Qy	307	TTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTGATGGCAAGCCTCCTGATATGAAG	366
Db	319	TTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTGATGGCAAGCCTCCTGATATGAAG	378
Qy	367	AAACAAGAGCTTGCTAAAGATCTCAAAAAGAGATGATGCAACCAAGATCTGACTGAG	426
Db	379	AAACAAGAACTTGCTAAAGATCTCAAAAAGAGATGATGCAACCAAGATCTGACTGAG	438

```

RESULT 3
US-09-805-311-3
; Sequence 3, Application US/09805311
; Patent No. US20010034886A1
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Rad2/FEN-1 Orthologues and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0961D
; CURRENT APPLICATION NUMBER: US/09/805,311
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 09/426,557
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/112,332
; PRIOR FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1541
; TYPE: DNA
; ORGANISM: zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (79)...(1215)
US-09-805-311-3

Query Match          96.4%; Score 1409.8; DB 10; Length 1541;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1411; Conservative 0; Mismatches 2; Indels 0; Gaps

QY 19  CCGCGGTTCTTCGGCCACTCCGGCTCAGCGCGCGCCGCCACCGCCACAGCCGCCCGCA 78
    |
DB 13  CCGCGGTTCTTCGGCCACTCCGGCTCAGCGCGCGCCGCCACCGCCACAGCCGCCCGCA 72
    |
QY 79  GACGAGATGGSCATCAAGGGTTTGACGAAACTGCTGGCGGACAATGCGCCCCAAGCGCATG 138
    |
DB 73  GACGAGATGGSCATCAAGGGTTTGACGAAACTGCTGGCGGACAATGCGCCCCAAGCGCATG 132
    |
QY 139  AAGGAGCAGAAAGTTCGAGAGCTACTTCGGCGCCGAAATCGCGTCGACGCCACGATGAGC 198
    |
DB 133  AAGGAGCAGAAAGTTCGAGAGCTACTTCGGCGCCGAAATCGCGTCGACGCCACGATGAGC 192
    |
... QY 199  ATATACCAAGTTCCTGATTGTAGTTGGAAGGACAGGCATGSAAACTCTCACAAATGAAGCT 258
    |
DB 193  ATATACCAAGTTCCTGATTGTAGTTGGAAGGACAGGCATGSAAACTCTCACAAATGAAGCT 252
    |
QY 259  GGTGAAGTCATAGTCATTTGCRANGAATGTTCAACCGGACATAAGATTACTGGAAGCG 318
    |
DB 253  GGTGAAGTCATAGTCATTTGCRANGAATGTTCAACCGGACATAAGATTACTGGAAGCG 312
    |
QY 319  GGAATCAAGCCAGTTTATGTTTTGATGGCAAGCCTCCTGATGTAAGAAACAAGAGCTT 378
    |
DB 313  GGAATCAAGCCAGTTTATGTTTTGATGGCAAGCCTCCTGATGTAAGAAACAAGAGCTT 372
    |
QY 379  GCTAAAGATACTCAAAAGAGATGATGCAACCAAGATCTGACTGNGCAGCTAGAGGTA 438
    |
DB 373  GCTAAAGATACTCAAAAGAGATGATGCAACCAAGATCTGACTGAGGCAAGTAGAGGTA 432
    |
QY 439  GCGAGATAAAGATGCGATTGAAAAATTGAGCAAGAGGACTGTAAGGTTCACAAAGCAACAC 498
    |
DB 433  GCGAGATAAAGATGCGATTGAAAAATTGAGCAAGAGGACTGTAAGGTTCACAAAGCAACAC 492
    |
QY 499  AACGAAGATGTAAACGGCTATTAAAGACTTTATGGGGTTTCCTGTTGTAGAGGCAACCTTCT 558
    |
DB 493  AACGAAGATGTAAACGGCTATTAAAGACTTTATGGGGTTTCCTGTTGTAGAGGCAACCTTCT 552
    |
QY 559  GAAGCAGAGCAGAGATGTGCAGCCCTTTTGCAATAACGATGAAGGTGTGCGCTGTTGCTTCA 618
    |
DB 553  GAAGCAGAGCAGAGATGTGCAGCCCTTTTGCAATAACGATGAAGGTGTGCGCTGTTGCTTCA 612
    |
QY 619  GAAGATATGGACTCCCTTACTTTTGGGGCTCCACGGTTCCTGTCATTTAATGGATCCA 678
    |

```

```
Db 613 GAAGATATGGACTCCCTTACTTTTGGGCTCCAGGFTCCCTCGTCATTTAATGGATCCA 672
Qy 679 AGTTCCAGAAATACCTGTGATGAATTTGATGTTGCCAAGGTTTTGGAGGAGCTTGAA 738
Db 673 AGTTCCAGAAATACCTGTGATGAATTTGATGTTGCCAAGGTTTTGGAGGAGCTTGAA 732
Qy 739 CTCACATGGACCAAGTTCATTTGATTTGTCATCTCGTGTGATGTGACTATTTGTGATAGC 798
Db 733 CTCACATGGACCAAGTTCATTTGATTTGTCATCTCGTGTGATGTGACTATTTGTGATAGC 792
Qy 799 ATCAAGATATCGGGGGGCAACAGCTCTGAAACTTATTCGTCAACATGGGTCCATAGAA 858
Db 793 ATCAAGATATCGGGGGGCAACAGCTCTGAACTTATTCGTCAACATGGGTCCATAGAA 852
Qy 859 AGCATCTTGGAGATCTTATAAAGACAGATATCAAAATTCCTGAGGACTGGCCTTACCAA 918
Db 853 AGCATCTTGGAGATCTTATAAAGACAGATATCAAAATTCCTGAGGACTGGCCTTACCAA 912
Qy 919 GAAGCTCGACGCTTGTTCAGAGGACCTAATGTACATTTGATATTCCTGAGCTAAAATGG 978
Db 913 GAAGCTCGACGCTTGTTCAGAGGACCTAATGTACATTTGATATTCCTGAGCTAAAATGG 972
Qy 979 ACTGACCTGATGAGGAGGCTCTCATTAAGTTTCCTGTGTAAGATTAATGTTTCAACGAA 1038
Db 973 ACTGACCTGATGAGGAGGCTCTCATTAAGTTTCCTGTGTAAGATTAATGTTTCAACGAA 1032
Qy 1039 GATCGGTTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGATAAATCGTCGCAAGGA 1098
Db 1033 GATCGGTTGAGAAAGGCCATAGAGAAGATCAAAATCTGCCAAGATAAATCGTCGCAAGGA 1092
Qy 1099 AGACTGAGTCTTTTCAAGCCAACTGCGCACCATCATCAGCACCGCTTAAACGGAAGGAG 1158
Db 1093 AGACTGAGTCTTTTCAAGCCAACTGCGCACCATCATCAGCACCGCTTAAACGGAAGGAG 1152
Qy 1159 ACTTCGGATAAAACAAGCGAGCTCGGAACAAGAAAAACAAGGCTGCTGGAAGAAG 1218
Db 1153 ACTTCGGATAAAACAAGCGAGCTCGGAACAAGAAAAACAAGGCTGCTGGAAGAAG 1212
Qy 1219 AAATAATCTTGATGCTTGAATACACTACGACTACGAAGCAGCGGTGGCGTGTATCAC 1278
Db 1213 AAATAATCTTGATGCTTGAATACACTACGACTACGAAGCAGCGGTGGCGTGTATCAC 1272
Qy 1279 TTGCGTTAGATATTTAACTCCCTGTTTTAACTCAGAGCTTTGGTAAAGCTTTGCTCATG 1338
Db 1273 TTGCGTTAGATATTTAACTCCCTGTTTTAACTCAGAGCTTTGGTAAAGCTTTGCTCATG 1332
Qy 1339 TTCAAGCTGGGGTAAAGTTAGTTGTTGTTGAAGAGATTTGGTACCAAGTAACAAAACCTT 1398
Db 1333 TTCAAGCTGGGGTAAAGTTAGTTGTTGTTGAAGAGATTTGGTACCAAGTAACAAAACCTT 1392
Qy 1399 ATCGCTGTTTTTACTTCTTCTGCTCTTTGAAAGTA 1431
Db 1393 ATCGCTGTTTTTACTTCTTCTGCTCTTTGAAAGTA 1425
```

## RESULT 4

```
US-09-805-311-5
; Sequence 5, Application US/09805311
; Patent No. US20010034886A1
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Rad2/FEN-1 Orthologues and Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 0961D
; CURRENT APPLICATION NUMBER: US/09/805,311
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 09/426,557
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/112,332
; PRIOR FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
```

```
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37)...(1173)
US-09-805-311-5
```

```
Query Match 92.3%; Score 1350.2; DB 10; Length 1381;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1355; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 49 CGCCGCCGCCACCCGCGCACAGCGAGATGGCATCAAGGTTTGACGAAA 108
Db 1 CGACCACACGCGTCCGCGCGCACAGCGAGATGGCATCAAGGTTTGACGAAA 60
Qy 109 CTGCTGGCGGCAATTCGCCCAAGCGCATGAAGGAGCAGAAAGTTGCGAGAGCTACTTCGGC 168
Db 61 CTGCTGGCGGCAATTCGCCCAAGCGCATGAAGGAGCAGAAAGTTGCGAGAGCTACTTCGGC 120
Qy 169 CGCAAAATCGCCGTCGAGCGCGCAGCATATACCATGCTCCCTGATTTGTTGGAAGG 228
Db 121 CGCAAAATCGCCGTCGAGCGCGCAGCATATACCATGCTCCCTGATTTGTTGGAAGG 180
Qy 229 ACAGGCGATGGAACCTCTCACAAATGAAGCTGGTGAAGTCACCTAGTCAATTTGCAAGGAATG 288
Db 181 ACAGGCGATGGAACCTCTCACAAATGAAGCTGGTGAAGTCACCTAGTCAATTTGCAAGGAATG 240
Qy 289 TTCACCCGACAAATGAAGTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTGTATGCGC 348
Db 241 TTCACCCGACAAATGAAGTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTGTATGCGC 300
Qy 349 AAGCCTCTCTGATATGAAGAAACAAGAGCTTGCTAAAAGATACTCAAAAAGAGATGATCA 408
Db 301 AAGCCTCTCTGATATGAAGAAACAAGAGCTTGCTAAAAGATACTCAAAAAGAGATGATCA 360
Qy 409 ACCAAGATCTCAGTGGGCGAGTAGAGTAGAGATAAGATGCGATTCGATTCGAAATTTGAGC 468
Db 361 ACCAAGATCTCAGTGGGCGAGTAGAGTAGAGATAAGATGCGATTCGATTCGAAATTTGAGC 420
Qy 469 AAGAGACTGTAAAGGTTCACAAGGCAACACACAGAGATTTGTAACCGCTATTAAGACTT 528
Db 421 AAGAGACTGTAAAGGTTCACAAGGCAACACACAGAGATTTGTAACCGCTATTAAGACTT 480
Qy 529 ATGGGGGTTCCCTGTTGTAGAGGCGACCTTCTGAGCAGAGCAGATGTGCGACCCCTTTGC 588
Db 481 ATGGGGGTTCCCTGTTGTAGAGGCGACCTTCTGAGCAGAGCAGATGTGCGACCCCTTTGC 540
Qy 589 ATAAACGATTAAGGTGTTGCTGTTCTCAGAAAGATATGGACTCCCTTACTTTTGGGGCT 648
Db 541 ATAAACGATTAAGGTGTTGCTGTTCTCAGAAAGATATGGACTCCCTTACTTTTGGGGCT 600
Qy 649 CCACGGTTCCTTCGTCATTTAATGGATCCAGTTCCAGAAATAACCTGTGATGAAATTT 708
Db 601 CCACGGTTCCTTCGTCATTTAATGGATCCAGTTCCAGAAATAACCTGTGATGAAATTT 660
Qy 709 GATGTGCCAAGGTTTTGGAGGAGCTTGAACCTCACCATTGGAGACTTCATTTGATTTGCG 768
Db 661 GATGTGCCAAGGTTTTGGAGGAGCTTGAACCTCACCATTGGAGACTTCATTTGATTTGCG 720
Qy 769 ATCCTGTGTGGATGTGACTATTTGATAGCATCAAAAGTATCGGGGGGCAACAGCTCTG 828
Db 721 ATCCTGTGTGGATGTGACTATTTGATAGCATCAAAAGTATCGGGGGGCAACAGCTCTG 780
Qy 829 AAACCTATTTCGTCACATGGGTCCATAGAAAGCATCTTGGAGATCTTAAATAAGACAGA 888
Db 781 AAACCTATTTCGTCACATGGGTCCATAGAAAGCATCTTGGAGATCTTAAATAAGACAGA 840
Qy 889 TATCAAAATTCCTGAGGACTGGCCTTTACCAAGAGCTGACGCTTGTTCAGAGAGCTTAAT 948
Db 841 TATCAAAATTCCTGAGGACTGGCCTTTACCAAGAGCTGACGCTTGTTCAGAGAGCTTAAT 900
Qy 949 GTCACATTTGGATATTCCTGAGCTAAAATGGACTGACCTGATGAGGAGGCTCTCATAGT 1008
```



Db 901 GTACATTGATATTCCTGAGCTAAATGGACTGCACCTGATGAGGAGGCTCTCATAGT 960  
QY 1009 TTCCTGGTAAAGATAATGGTTTCAACGAAGATCGGCTGACAAAGCCATAGAGAATC 1068  
Db 961 TTCCTGGTAAAGATAATGGTTTCAACGAAGATCGGCTGACAAAGCCATAGAGAATC 1020  
QY 1069 AAATCTGCCAAGAATAAATCGTCGAAGGAAGACTCGAGTCCCTTTTCAAGCCAACTGCC 1128  
Db 1021 AAATCTGCCAAGAATAAATCGTCGAAGGAAGACTCGAGTCCCTTTTCAAGCCAACTGCC 1080  
QY 1129 ACCACATCAGCACCGCTAAACGAAGGAGACTTCGGATAAAAACAAGCAGGCTGCC 1188  
Db 1081 ACCACATCAGCACCGCTAAACGAAGGAGACTTCGGATAAAAACAAGCAGGCTGCC 1140  
QY 1189 ACAAGAAACAAAGGCTGGTGAAGGAAGAAATAATCTTGGATGCTGTATGATACACTA 1248  
Db 1141 ACAAGAAACAAAGGCTGGTGAAGGAAGAAATAATCTTGGATGCTGTATGATACACTA 1200  
QY 1249 CGACTACGAAAGCAGCGGTGGCTGATCACTTCGCTTAGATTTATTAACCTCCCTGTTTTA 1308  
Db 1201 CGACTACGAAAGCAGCGGTGGCTGATCACTTCGCTTAGATTTATTAACCTCCCTGTTTTA 1260  
QY 1309 ACTCAGAGCTTTGGTAAAAATTTGCTCATGTTTCAAGCTGGGTAAGTTAGTTGTGTTG 1368  
Db 1261 ACTCAGAGCTTTGGTAAAAATTTGCTCATGTTTCAAGCTGGGTAAGTTAGTTGTGTTG 1320  
QY 1369 AAGAGATTGGTACCAGTAACAAACTTATCGCTGTTTTT 1411  
Db 1321 AAGAGATTGGTACCAGTAACAAACTTATCGCTGTTTTT 1363

## RESULT 5

US-10-033-297-78

; Sequence 78, Application US/10033297

; Publication No. US20020187486A1

; GENERAL INFORMATION:

; APPLICANT: Hall, Jeff G.

; Mast, Andrea L.

; Brow, Mary Ann D.

; TITLE OF INVENTION: Detection of Nucleic Acids By Multiple  
; Sequential Invasive Cleavages

; NUMBER OF SEQUENCES: 163

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Medlen &amp; Carroll, LLP

; STREET: 220 Montgomery Street, Suite 2200

; CITY: San Francisco

; STATE: California

; COUNTRY: United States Of America

; Zip: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/033.297

; FILING DATE: 12-NOV-2001

; CLASSIFICATION DATA:

; APPLICATION DATA: &lt;Unknown&gt;

; PRIORITY DATA:

; FILING DATE: 09/09/350.597

; APPLICATION NUMBER: US/08/823.516

; FILING DATE: 24-MAR-1997

; APPLICATION NUMBER: PCT/JUS97/01072

; FILING DATE: 21-JAN-1997

; APPLICATION NUMBER: US 08/759,038

; FILING DATE: 02-DEC-1996

; APPLICATION NUMBER: US 08/758,314

; FILING DATE: 02-DEC-1996

; APPLICATION NUMBER: US 08/756,386

; FILING DATE: 29-NOV-1996

APPLICATION NUMBER: US 08/682,853

FILING DATE: 12-JUL-1996

APPLICATION NUMBER: US 08/599,491

FILING DATE: 24-JAN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: FORS-02736

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:

LENGTH: 1023 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1020

SEQUENCE DESCRIPTION: SEQ ID NO: 78:

US-10-033-297-78

Query Match 10.9%; Score 159.6; DB 9; Length 1023;

Best Local Similarity 54.6%; Pred. No. 2.4e-34;

Matches 318; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 134 CGATCAAGGAGCAGAAGTTTCGAGAGCTACTTCGGCCGCAAAATCGCGTCGAGCGCCAGCA 193  
Db 29 CAAGAAAAAGAAATTGAGTTAGAAAAACCTATACGGGAAAAAAATCGCAATCGACGCTCTTA 88  
QY 194 TGAGCATATACCAAGTTCCTGATTGTAGTTGGAAGGACAGGCGATGGAACTCTCACAAATG 253  
Db 89 ATGCAATCTACCAATTTTGTCCACATAAGACAGAAAGATGCAACTCCACTTATGGATT 148  
QY 254 AAGCTGGTGAAGTCACTAGTCAATTTTGCAGGAATGTTCAACCGGAGCAATTAAGATTACTGG 313  
Db 149 CAAAGGTAGATAAATACCTCCACCTAAAGGGGCTCTTTTACAGGACAAATAACCTAATGG 208  
QY 314 AAGCGGGAATCAAGCCAGTTTATGTTTGTATGCGCAAGCTCTCTGATATCAAGAACAAG 373  
Db 209 AGGCTGGAAATAAACCTGTGTATGTTTGTATGGAACCTCCAGATTCCAAAAGAAAG 268  
QY 374 AGCTTGCTAAAGAGATACTCAAAAGAGATGATGCAACCAAGATCTGACTCAGGCACTAG 433  
Db 269 AGCTCGAAAAAGAGAGAGAGCGAGAGAGAACTGAAGAAAAAGTGGAGAGAGCACTTGG 328  
QY 434 AGGTAGGAGATAAAGATGCGATTGAAAAATTTGAGCAAGAGGACTGTAAAGTCAAGGC 493  
Db 329 AAAAAAGGAGATAGAGGAAGCAAGAAAAATATGCCCAAGAGCAACCCAGGTAATGAAA 388  
QY 494 AACACAACGAAGATTGTAACGGCTATTAGACTTTATGGGGTTTCCCTGTTGTAGAGCAC 553  
Db 389 TGCTCATCGAGGATGCAAAAAAATCTTTAGAGCTTTATGGGAATTCCTATAGTTCAGCAC 448  
QY 554 CTCTTGAAGCAGAAGCAGAAATGTGCAGCCCTTTTGCATATAAAGCAAGGTTTCCTGTTG 613  
Db 449 CTAGCGAGGAGAGGCCCAAGCTGATATATGCGCAAGAGGGAGCGTGTATGCATCGG 508  
QY 614 CTTCAAGAGATATGGACTCCCTTACTTTTGGGGCTCCACGGTTCCTTCGTCATTTATGG 673  
Db 509 CTAGTCAAGATTACGATTCCCTACTTTTGGAGCTCCAGACTTGTGTAGAACTTAACAA 568  
QY 674 ATCCCAAGTTCCAAGAAATACCTGTGATGCAATTTGATGTTG 715  
Db 569 TAACAGGAAAAAGAAAGTTGCCCTGGGAAAAAATGCTACGTCG 610

## RESULT 6

US-09-940-244-78

; Sequence 78, Application US/09940244

; Publication No. US20030044796A1

## ; GENERAL INFORMATION:

; APPLICANT: Neri, Bruce P.  
; APPLICANT: Hall, Jeff G.  
; APPLICANT: Lyamichiev, Victor  
; APPLICANT: Smith, Lloyd M.  
; TITLE OF INVENTION: Reactions on Dendrimers  
; FILE REFERENCE: FORS-06478  
; CURRENT APPLICATION NUMBER: US/09/940,244  
; CURRENT FILING DATE: 2002-05-06  
; NUMBER OF SEQ ID NOS: 422  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 78  
; LENGTH: 1023  
; TYPE: DNA  
; ORGANISM: Pyrococcus furiosus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1020)  
; OTHER INFORMATION:  
US-09-940-244-78

Query Match 10.9%; Score 159.6; DB 9; Length 1023;  
Best Local Similarity 54.6%; Pred. No. 2.4e-34;  
Matches 318; Conservative 0; Mismatches 264; Indels 0; Gaps 0;  
QY 134 CGATGAGGAGCAGAGTTCGAGAGCTACTTCGGCCGCCAAATTCGCGTCGAGCCAGCA 193  
Db 29 CAAGAAAGAAATTTGATTTAGAAAACCTATACGGGAAAAAATTCGCAATCGAGCTCTTA 88  
QY 194 TGAGCATATACCAAGTTCCTCATTTGTTGGAAGGAGCAGCATGGAACCTTCACAATG 253  
Db 89 ATGCATCTACCAATTTTGTCCACATAAAGACAGAAGATGGAATCCACCTTATGGATT 148  
QY 254 AAGCTGGTGAAGTCACTAGTCAATTTGCAAGGAATTTCAACCGGACAAATTAAGATTACTGG 313  
Db 149 CAAAGGAGTAAATTAACCTCCACCTTAAGCGGCTCTTTTACAGGACAAATTAACCTTAATGG 208  
QY 314 AAGCGGGAATCAAGCAGTTTATGTTTTTGTGATGCGAAGCTCTCTGATATGAAGAAACAAG 373  
Db 209 AGCTGGAATTAACCTGTGTATGTTTTTGTGATGGAACCTCCAGAAATTCAGAAAGAAAG 268  
QY 374 AGCTTGCTAAAGATACTCAAAAGAGATGATGCAACCAAGATCTGACTGAGGCGAGTAG 433  
Db 269 AGCTCGAAAAG 328  
QY 434 AGTGAAGATAAAGATGCGATTGAAAAATTTGACGAAGAGGACTGTAAAGGTCAACAAGGC 493  
Db 329 AAAAGGAGAGATAGAGGAAGCAAGAAATATATGCCCAAGAGCAACACCGGTAATGAA 388  
QY 494 AACACAACAGAGATTTGAACGCTATTAAAGCTTTAGCAAGAGGACTGTAAAGGTCAACAAGGC 553  
Db 389 TGCTCATCGAGGATGCAAAAAAATCTTATAGAGCTTTATGGGAATTCCTATAGTTCAAGCAC 448  
QY 554 CTCTCTGAAGCAGAGCAGATGTCAGCCCTTTGCAATAAAGCATAAGGTGTCGCTGTGG 613  
Db 449 CTAGCAGGAGAGAGGCGCAAGCTGATATATGCCCAAGAGGAGGCGGTGATGATCGG 508  
QY 614 CTTCAAGAGATATGAGTCCCTTACTTTTGGGGCTCCACGGTTCCCTGCTCAATTTAATGG 673  
Db 509 CTAGTCAAGATTACGATTCCTTACTTTTGGAGCTCCAAAGACTTTGTAGAACTTAACAA 568  
QY 674 ATCCAAGTTCCAGAAATACCTGTGATGGAATTTGATGTTG 715  
Db 569 TAACAGAAAAAGAAAGTTGCTCGGAAAAAATGCTACGTCG 610

## RESULT 7

US-09-940-244-281  
; Sequence 281, Application US/09940244  
; Publication No. US20030044796A1  
; GENERAL INFORMATION:  
; APPLICANT: Neri, Bruce P.  
; APPLICANT: Hall, Jeff G.

; APPLICANT: Lyamichiev, Victor  
; APPLICANT: Smith, Lloyd M.  
; TITLE OF INVENTION: Reactions on Dendrimers  
; FILE REFERENCE: FORS-06478  
; CURRENT APPLICATION NUMBER: US/09/940,244  
; CURRENT FILING DATE: 2002-05-06  
; NUMBER OF SEQ ID NOS: 422  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 281  
; LENGTH: 1054  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-940-244-281

Query Match 10.4%; Score 152.8; DB 9; Length 1054;  
Best Local Similarity 53.4%; Pred. No. 2e-32;  
Matches 350; Conservative 0; Mismatches 297; Indels 9; Gaps 1;  
QY 134 CGATGAGGAGCAGAGTTCGAGAGCTACTTCGGCCGCCAAATTCGCGTCGAGCCAGCA 193  
Db 73 CAAGAAAGAAATTTGATTTAGAAAACCTATACGGGAAAAAATTCGCAATCGAGCTCTTA 132  
QY 194 TGAGCATATACCAAGTTCCTCATTTGTTGGAAGGAGCAGCATGGAACCTTCACAATG 253  
Db 133 ATGCAATCTACCAATTTTGTCCACATAAAGACAGAAGATGGAACCTCCACTTATGGATT 192  
QY 254 AAGCTGGTGAAGTCACTAGTCAATTTGCAAGGAATTTCAACCGGACAAATTAAGATTACTGG 313  
Db 193 CAAAGGAGTAAATTAACCTCCACCTTAAGCGGCTCTTTTACAGGACAAATTAACCTAATGG 252  
QY 314 AAGCGGGAATCAAGCAGTTTATGTTTTTGTGCAAGCTCTCTGATATCAAGAAACAAG 373  
Db 253 AGCTGGAATTAACCTGTGTATGTTTTTGTGATGAGNACCTCCAGATTCACAAAGAAAG 312  
QY 374 AGTTGCTAAAGATACTCAAAAGAGATGATGCAACCAAGATCTGACTGAGGCGAGTAG 433  
Db 313 AGCTGAAAAAG 372  
QY 434 AGCTAGGAGATAAGATGCGATTGAAAAATTTGACGAAGAGGACTGTAAAGGTCAACAAGGC 493  
Db 373 AAAAGGAGAGATAGAGAAAG 432  
QY 494 AACACAACAGAGATTTAAACGCTATTAAAGCTTTATGGGGTTTCTGTTGTAGAGGAC 553  
Db 433 AATGCTTGAACACTCAATATTTGTTAAGTTTGTAGTGGCATTCGCTATGTTGAAGCTC 492  
QY 554 CTTCTGAAGCAGAGCAGATGTCAGCCCTTTGCAATAAAGCATAAGGTGTCGCTGTGG 613  
Db 493 CCTCTGAGGAGAGGACACAAAGCAAGCTATATGCAAAAGAGGAGAGATGTTGGGAGTTG 552  
QY 614 CTTCAAGAGATATGAGTCCCTTACTTTTGGGGCTCCACGGTTCCCTGCTCAATTTAATGG 673  
Db 553 TAAGTCAAGATTTATGCTCTGTTTATATGGAGCTCCGAGAGTTGTAGAAATTTAACAA 612  
QY 674 ATCCAAGTTCCAGAAATACCTGTGATGGAATTTGATGTTGCCAAGGTTTTCGAGGAGC 733  
Db 613 CTACAAAGAGAGATGCCAGAACTTA-----TTCAATTAATGAGGTTTATAGAGATT 663  
QY 734 TTGAAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 789  
Db 664 TAAGAAATTTCTTGGATGATTTGATAGATATAGCCATATATTTATGGGAAGTACTAT 719

## RESULT 8

US-09-918-995-30803  
; Sequence 30803, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES

```

; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30803
; LENGTH: 485
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(485)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30803

```

```

Query Match      10.3%; Score 151.4; DB 9; Length 485;
Best Local Similarity 61.4%; Pred. No. 2.9e-32;
Matches 261; Conservative 0; Mismatches 161; Indels 3; Gaps 1;

QY 107 AACTGCTGCGGACATCGCCCGGAGGATGAAGGAGCAGAGTTCGAGAGCTACTTCG 166
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64 AACTAATTGCTGATGTGGCCCGCCAGTCCATCGGGAGAGATGACATCAAGAGCTACTTTG 123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 167 GCGGAAAAATCGCGGTCGACGCGCAGCATGAGCATATACCATTCCTGTGATTTGGAA 226
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124 GCGTAAAGGTGGCCATTGATGCTCTATGAGCATTTATCAGTTCCCTGATTCGTCGCC 183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 227 GGACAGCATGTGAACCTTCACAAATGAAGCTGGTGAAGTCACTAGTCAATTTGCAAGGAA 286
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 AG---GGTGGGGATGTGCTGCAGAAATGAGAGGGTGAGACACCCAGCCACTGATGGCA 240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 287 TGTTCAACGCGCAATAAGATTACTGSAAGCGGGAATCAAGCCAGTTTATGTTTTGATG 346
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TGTTTACCGCCACCATTCGCGATGATGAGAACGCGCATCAAGCCCGTGTATGCTTTGATG 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 347 GCAAGCCTCTGTATGAAGAAACAAGAGCTTGCTTAAAGATATCTCAAAAGAGATGATG 406
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GCAAGCGCCACAGCTCAAGTCAGGCGAGCTGGCCAAACGACGTGAGCGCGGCTGAGG 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 407 CAACCAAGATCTGACTGAGGAGTAGAGTAGAGATAAGATGCGATTGAATAATTGA 466
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CAGAGAAGAGCTGCGACAGCTGCTGAGGCTGCTGGGCGGAGCAGAGGTGGAATAATTCA 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 467 GCAAGAGAGCTGTAAGGTCAACAGGCAACACAGCAAGATTTGTAACGGCTATTAGAC 526
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 CTAAGCGGTGCTCAAGGTCACTAAGCAGCACAATGATGATGATGCAACATCTGCTGAGCC 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 527 TTATG 531
      |||||
Db 481 TCATG 485

```

```

RESULT 9
US-09-940-244-289
; Sequence 289, Application US/09940244
; Publication No. US20030044796A1
; GENERAL INFORMATION:
; APPLICANT: Neri, Bruce P.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor
; APPLICANT: Smith, Lloyd M.
; TITLE OF INVENTION: Reactions on Dendrimers
; FILE REFERENCE: FORS-06478
; CURRENT APPLICATION NUMBER: US/09/940,244
; CURRENT FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 289
; LENGTH: 1115
; TYPE: DNA
; ORGANISM: Artificial Sequence

```

```

; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-940-244-289

```

```

Query Match      10.2%; Score 149.6; DB 9; Length 1115;
Best Local Similarity 53.0%; Pred. No. 1.6e-31;
Matches 348; Conservative 0; Mismatches 299; Indels 9; Gaps 1;

QY 134 CGATGAAGGAGCAGAGTTTCGAGAGCTACTTCGCGCGCAAAATCGCGTCGAGCG/CAGCA 193
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73 CAAGAAAAGAAATAGTTAGAAAACCTATACGGGAAAAAATTCGAATCGAGCTCTTA 132
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 194 TGAGCATATACAGTTCCTGTAGTTGTAGTTGGAAGGACAGCATGGAACCTCTCACAATG 253
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 ATGCATCTACCAATTTTGTCCACAATAAGACAGAAAGATGGAATCCACATTCATGGATT 192
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 254 AAGCTGTGAAGTCACCTAGTCTATTTCAGAGGAATGTTCAACCGGACAATAAGATTTACTGG 313
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193 CAAAGGTAGATACTCTCCACCTAAGCGGCTCTTTTACAGGACATAAACCCTAATGG 252
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 314 AAGCGGGAATCAAGCCAGTTTATGTTTGTATGGAAGCCTCTGTATATGAAGAAACAAG 373
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 253 AGSCTGGAATAAAACCTGTGTATGTTTGTATGATGAGAACCCTCCAGAATTCAAAAGAAAG 312
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 374 AGCTTCTTAAAGATACTCAAAAAGAGATGATGCARCCCAAGATCTCAGCTGAGCAGTAG 433
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 313 AGCTCGAAAAGAGAGAGAGAGCGAGAGAGAGCTGAAAGAAAAGTGGAGAGAGCAGCTTG 372
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 434 AGTAGGAGATAAGATGCGATTGAAAATTTAGCAAGAGGAGCTGTAAGAGTCAACAGGC 493
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 373 AAAAGAGAGATAGAGAGCAAGAAATATGCCCAAGAGCAACAGGTAATGAAA 432
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 494 AACCAACGAGATGTTTAAACGGCTATTAAAGACTTATGGGGTTCCTGTGTAGAGCAC 553
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 433 TGCTCATCGAGGATGCAAAAATCTCTTAGAGCTTATGGGAATTCCTATATAGTTCAGCAC 492
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 554 CTCTCAGCAGAGAGAGATGTGCAGCCCTTTGCATAAAGCAAGTGTGCTGCTGTTG 613
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 493 CTAGCGGAGGAGGCGCCCAAGCAAGCTATATGCAAGAGAGGAGAGTGTGTCGAGTTG 552
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 614 CTTTCAGAGATATGAGCTCCCTTACTTTTGGGCTCCACGCTTCCTTCGTCATTTAATGG 673
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 553 TAAGTCAAGATTATGATGCTTGTATATGAGCTCCGAGAGTTGTTAGAAATTTAACAA 612
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 674 ATCCAAAGTCCAGAAAATACCTGTGATGGAATTTGATGTTGCCAAGTTTGGAGGAGC 733
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 613 CTACAAAGGAGATGCCAGAACTTA-----TTGAATTAAGTGTGTTAGAGGATT 663
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 734 TTGAAGTCAAGTGGAGCAGTTCATTTGTCATCTGTCCTGTGTGATGATGACAT 789
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 664 TAAGATTTCTTTGGATGTTTATAGATATAGCCATATTATGGGAACATGACAT 719
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 10
US-09-864-761-16172/c
; Sequence 16172, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03

```

; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 15172  
 ; LENGTH: 510  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC013557.3  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2  
 US-09-864-761-16172

Query Match 9.5%; Score 139.6; DB 10; Length 510;  
 Best Local Similarity 60.1%; Pred. No. 6.3e-29;  
 Matches 268; Conservative 0; Mismatches 174; Indels 4; Gaps 2;  
 QY 85 ATGGGATCAAGGTTTGACGAACACTGCTGGCGGACAAATCGCCCAAGCGGATGAAGGAG 144  
 DB 443 ATGGGATTTCAAGGCTGGCCAACTAAATGCTGATGTGCGCCCGCAGTGCACCCGGGAG 384  
 QY 145 CAGAAGTTCGAGAGTACTTCGCGCGCAAAATCGCGTCGACGCCAGCATGACATATAC 204  
 DB 383 AATGACATCAAGAGTACTTTGGCCGTAGGTGGCCATTGATGCTCTATGAGCACTAT 324  
 QY 205 CAGTCTCTGATTTAGTGTGAAGGACAGGATGGAACCTCTCACAATGAAGCTGTGAA 264  
 DB 323 CAGTCTCTGATTTGTCGCCAG---GGTGGGATGCTGTCGAGACAGGAGGTGAG 267  
 QY 265 GTCACTAGTCAATTTGAAGATTTTCAACCGGACAAATAGATTACTTGAAGCGGGAATC 324  
 DB 266 ACCACGAGCCACCTGTAGGATGATTTCTACCGCACTATTGCGGTGATGGAGAACGGCATC 207  
 QY 325 AAGCCAGTTTATTTTTGATGGCAAGCTCTCTGATATGAAGAAACAGAGCTTGCTAA 384  
 DB 206 AAGCCATGATGCTTTGAGGCGACGCCACACACAGCAAGTCAGCGGCTGGGCCAA 147  
 QY 385 AG-ATACATAAAGAGATGATCAACCAAGATCTGACTTGAGCGAGTAGAGAGTAGAGA 443  
 DB 146 TGCAGTGAACCTGCGGGCTGAGGACAGAGAGCAGCTACAGAGGCTCAGGCTGCTGGGC 87  
 QY 444 TAAAGATGCGATTGAAAATTTAGGCAAGAGGACTGTAAAGGTCACAAAGGCACACAAACA 503  
 DB 86 CGAGCAGGTGGTGGAAAAATTCATAAGTGGTGGTGAAGGTCACTTAAGCAGCACAAACA 27

QY 504 AGATTGTAAACGGCTATTAAAGACTTA 529  
 DB 26 TAAGTGAACATCTGCTGAGCCTCA 1  
 RESULT 11  
 US-09-940-244-356  
 ; Sequence 356, Application US/09940244  
 ; Publication No. US2003004796A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Neri, Bruce P.  
 ; APPLICANT: Hall, Jeff G.  
 ; APPLICANT: Lyamichov, Victor  
 ; APPLICANT: Smith, Lloyd M.  
 ; TITLE OF INVENTION: Reactions on Dendrimers  
 ; FILE REFERENCE: FORS-06478  
 ; CURRENT APPLICATION NUMBER: US/09/940,244  
 ; CURRENT FILING DATE: 2002-05-06  
 ; NUMBER OF SEQ ID NOS: 422  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 356  
 ; LENGTH: 1032  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Synthetic  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1032)  
 ; OTHER INFORMATION:  
 ; US-09-940-244-356

Query Match 9.3%; Score 135.6; DB 9; Length 1032;  
 Best Local Similarity 52.1%; Pred. No. 1.3e-27;  
 Matches 303; Conservative 0; Mismatches 279; Indels 0; Gaps 0;  
 QY 134 CGATGAAGGACGAGAGTTCGAGAGTACTTCGCGCCCAAAATCGCGCTCAGCCAGCA 193  
 DB 29 CGAGGAGGAGATAGATCTTGAATACTGTATGGAAGAAGATAGCGATAGATGCGCCTAA 88  
 QY 194 TGAGCATATACCATCTCTGATTTGGAAGGACAGCATGGAACACTCTCACAATG 253  
 DB 89 ACCCATCTATCATGTTTTTATCAACGATAGACAGAGGATGGAACACCATCTATGGACT 148  
 QY 254 AAGCTGTGAAGTCACTAGTCTATTTCAGAGGAATTTCAACCGGACAAATAGATTACTGG 313  
 DB 149 CTAAGGTTAGGATAACCTCTCATTTAAGTGGCTCTTTTATAGACGATAAATCTAATGG 208  
 QY 314 AAGCGGAATCAAGCCAGTTTATGTTTTGATGGCAAGCTCTCTGATATGAAGAAACAAG 373  
 DB 209 AAGCCGTTATTAAGCGCGGCTACGCTCTTTGATGAAAGCTCCGGAATTCAAAAGGAAG 268  
 QY 374 AGCTTCTAAAGATACTCAAAAAGAGATGATCAACCAAGATCTCACTGAGCGCAGTAG 433  
 DB 269 AGCTCAAAAAGAGGAGAGCTAGAGAGAGCGCACTAAATGGAAGAAGCTCTAG 328  
 QY 434 AGGTAGGATTAAGATGCGATTGAAAAATTTGAGCAAGAGGACTGTAAAGGTCACAAAGGC 493  
 DB 329 CCAAGGAAACCTGGAGGAGCTAGGAAATACGCTCAAGGGCAACTAAGGTTAATGAAA 388  
 QY 494 AACACAAAGAGATTCTAAACGCTATTAAAGACTTATGGGGGTTCTGTTGTAGAGGCAC 553  
 DB 389 TGCTAATCAAGAGATGCAAGAAGCTTTTGCACTAATGGAATACCAATAAATTCAGGCTC 448  
 QY 554 CTCTCTCAAGCAGAGCAGATGTCAGCCCTTTTGCATAAAGGATAGGTTGCGCTGTTG 613  
 DB 449 CAAAGTGAAGAGAGAGCCCAAGCGCATACATGGAAGTAAAGGGGATGCTACGCGCTCAG 508  
 QY 614 CTTCAGAGATATGGAATCCCTTACTTTTGGGCTCCAGCGTTCCCTTCGCTCAATTAATGG 673  
 DB 509 CGAGTCAAGATTATGATTCACCTACTCTTTGGTGGTCCCAAGGTTGATTAGGAATCTGACAA 568



Db 313 AGCTGAAAAAAGAGAGAGGAGAGAGGAGAGAGTGAAGAAAAAGTGGAGAGAACACTTG 372  
QY 434 AGGTAGGAGATAAGATGCGATTGAAAAATTGACGAAGAGGACTGTAAAGGTCAAAAGGC 493  
Db 373 AAAAAGGAGAGATAGAGGAAGCAAGAAATATGCCAAAGAGCAACACAGGTAAATGAAA 432  
QY 494 AACACACGAGAGATTGAAACGGCTATTAAGACTTATGGGGGTTCCGTGTTGTAAGAGCAC 553  
Db 433 TGCTATCGAGGATGCAAAAAAATCTTAGAGCTTATGGGAATTCCTATAGTTCAGAGCAC 492  
QY 554 CTTCTGAAGCAGAGCAGAA 573  
Db 493 CTACGGAGGAGAGGCCCAA 512

RESULT 14  
US-09-940-244-383  
; Sequence 383, Application US/09940244  
; Publication No. US20030044796A1  
; GENERAL INFORMATION:  
; APPLICANT: Neri, Bruce P.  
; APPLICANT: Hall, Jeff G.  
; APPLICANT: Lyamichev, Victor  
; APPLICANT: Smith, Lloyd M.  
; TITLE OF INVENTION: Reactions on Dendrimers  
; FILE REFERENCE: FORS-06478  
; CURRENT APPLICATION NUMBER: US/09/940,244  
; CURRENT FILING DATE: 2002-05-06  
; NUMBER OF SEQ ID NOS: 422  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 383  
; LENGTH: 1023  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
; NAME/KEY: CDS  
; LOCATION: (1)..(1023)  
; OTHER INFORMATION:  
US-09-940-244-383

Query Match 8.7%; Score 127.6; DB 9; Length 1023;  
Best Local Similarity 51.2%; Pred. No. 2.3e-25;  
Matches 298; Conservative 0; Mismatches 284; Indels 0; Gaps 0;  
QY 134 CGATGAGGAGCAGAGTTCGAGAGCTACTTCGCGCGCAAAATCGCGTCGAGCGCCAGCA 193  
Db 29 CAAGAAAGAGCTTGAGCTTGAAATTAATGGGAGAAAGTTGGATAGATGCATTTA 88  
QY 194 TGAGCATATACCAAGTTCCTGATTGTAGTTGGAAGGACAGGCATGGAACCTCTCACAAATG 253  
Db 89 ACCTATTATACCAAGTTCTCTCAACATAAGACAAGAGATGGGACTCCTTTAATGGATT 148  
QY 254 AAGCTGTGAAGTCACTAGTCAATTCGACGAATGTTCAACCGGACAAATGAAGATTACTGG 313  
Db 149 CCAAGGAGAGATAAGCTCCATCTTCAGGGCTTTTACAGGACTATAAACCTAATGG 208  
QY 314 AAGCGGAATCAAGCCAGTTTATGTTTATGCGCAAGCTCCTGATATGAAGAAACAAG 373  
Db 209 AAGCGGAATAAAGCTGCGTATGTATTCGATGGGAAGCTCCAGAGTTCAAGAAAGAAG 268  
QY 374 AGCTTGTAAAGATACTCAAAAGAGATGATCCACCAAGATCTGACTGAGGAGTAG 433  
Db 269 AGCTTGAAGAAAGCTAGGAGCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 328  
QY 434 AGGTAGGAGATAAGATGCGATTGAAAAATTGAGCAGAGGAGTGTAAAGGTCAACAGGC 493  
Db 329 CAAGGGAGACTTAGAAGAGCGCAAGAAATATGCCAGCGGGCGGCAAGATAAATGAGA 388  
QY 494 AACACACGAGAGATTGAACGCTTATTAAGACTTATGGGGGTTCCGTGTTGTAAGAGCAC 553  
Db 389 TGCTTATCGAGGATGTAAGAGCTTTTGAGGCTTATGGGATCCCATGGGTGAGGCTC 448

QY 554 CTTCTGAAGCAGAGCAGAAATGTCAGCCCTTTGCATAAAACGATAAGGTGTTGCTGTTG 613  
Db 449 CTAGCGAAGTGAAGCGCAGCAGCTTATATGGCATCTAAAGGCGACGTTTGGCCCTCGG 508  
QY 614 CTTCAAGAAGATATGAGTCCCTTACTTTTGGGGTCCACCGGTTCCCTTCGTCATTAAATGG 673  
Db 509 CGAGCCAGGAGTACGACTCGCTCCTCTTCGGAACACCAAGGCTAGTGAGAAACCTCACCA 568  
QY 674 ATCAAGTTCCAAGAAATACCTCTGATGAATTTGATGTTG 715  
Db 569 TAACTGGAAGAGAAAGCTTCCTGGGAAGGATATTTACGTAG 610

RESULT 15  
US-09-940-244-336  
; Sequence 336, Application US/09940244  
; Publication No. US20030044796A1  
; GENERAL INFORMATION:  
; APPLICANT: Neri, Bruce P.  
; APPLICANT: Hall, Jeff G.  
; APPLICANT: Lyamichev, Victor  
; APPLICANT: Smith, Lloyd M.  
; TITLE OF INVENTION: Reactions on Dendrimers  
; FILE REFERENCE: FORS-06478  
; CURRENT APPLICATION NUMBER: US/09/940,244  
; CURRENT FILING DATE: 2002-05-06  
; NUMBER OF SEQ ID NOS: 422  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 336  
; LENGTH: 1056  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
; NAME/KEY: CDS  
; LOCATION: (1)..(1056)  
; OTHER INFORMATION:  
US-09-940-244-336

Query Match 8.6%; Score 126.4; DB 9; Length 1056;  
Best Local Similarity 52.0%; Pred. No. 5.1e-25;  
Matches 283; Conservative 0; Mismatches 261; Indels 0; Gaps 0;  
QY 172 AAAATCGCCCTCGACGCGCAGCATATACCACTTCCTGATTGTAGTTGGAAGGACA 231  
Db 73 AAAGTAAGCATAGATGCTTATACGCTATTTACCAGTTTTTGTACTGCAATAAGACAGCCA 132  
QY 232 GGCAATGGAACCTCTCAAAATGAAGCTGGTGAAGTCACTAGTCAATTTGCAAGGAATGTTT 291  
Db 133 GATGGTACTCCACTAATGGACTCACAAGGAAGAGTTACTAGTCATCTTAGTGAATATTT 192  
QY 292 AACCGGACATATAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTGTATGCAAG 351  
Db 193 TATAGAACATATAGCCCTTTTAGAAGAGGAGTAAATTCCTCAATTTATGTTATCGATGGA 252  
QY 352 CCTCTGATATGAAGAAACAAGAGCTTGTCTAAAGATCTCAAAAAGAGATGATGCAACC 411  
Db 253 CCACCAAGACTTAAAGCTCAAGATTTAGAAAGAAAGAAATAAAGGAAGAGCTGAG 312  
QY 412 AAAGATCTGACTGAGGAGTAGAGTAGAGATAAAGATGCGATTGAAAAATTTGAGCAAG 471  
Db 313 AAAAAATTTGAAAAGCCAAAGAAAGGAGGAGAAACAAAGGAATTAAGAAAGTATTCGCA 372  
QY 472 AGGACTGTAAAGGTCAACAGGCAACACACGGAAGATTGTAACGGCTATTAAAGCTTATG 531  
Db 373 ATGGCAACTAGGTTAATATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 432  
QY 532 GGGTTTCTCTTTGAGAGGACCTTCTTGAAGCAGAGAGAGAGATGTGCAGCCCTTTTGCATA 591  
Db 433 GGAATTCAGTAGTGCAGCTCAAGCTCAAGTGAAGGAGAGCTGAGCGCAGCGTATTTATGAGT 492







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 19:48:32 ; Search time 1963 Seconds

(without alignments)  
12070.299 Million cell updates/sec

Title: US-09-805-311-1

Perfect score: 1463  
Sequence: 1 caccagataatgctgcgccc.....aaaaaaaaaaaaaaaaaaaa 1463

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estlin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1425.8	97.5	1552	11	AY104316 Zea mays
2	778.8	53.2	901	12	BG837708 Zm10_01f0
3	614.8	42.0	638	14	BU037559 946140C05
4	560.8	38.3	586	9	AI881599 606068G09
5	536.4	36.7	550	10	BE639422 946033A02
6	534.4	36.5	554	10	BE639421 946033A02

7	517.8	35.4	553	10	AW562789
8	501	34.2	532	10	AW000375
9	465	31.8	648	10	AV913663
10	463.4	31.7	474	10	AW559173
11	462	31.6	470	9	AI861468
12	456.4	31.2	735	14	BQ986894
13	448	30.6	467	10	BE186786
14	429.8	29.4	475	10	AW562517
15	423.6	29.0	455	13	EM501417
16	420.4	28.7	733	10	AW774700
17	413	28.2	553	9	AI834484
18	411.6	28.1	539	12	BG263233
19	411	28.1	540	9	AJ469880
20	411	28.1	540	9	AJ469882
21	402.2	27.5	414	10	AW288784
22	400.8	27.4	540	9	AJ469881
23	398	27.2	456	9	AI065689
24	386.2	26.4	548	13	BJ211325
25	378.8	25.9	632	10	AW585913
26	373.6	25.5	653	14	BQ407057
27	372.6	25.5	609	12	BG096545
28	370.4	25.3	754	10	BE416584
29	369.6	25.3	820	12	BG414505
30	366.4	25.0	480	9	AJ467955
31	366.4	25.0	480	9	AJ468955
32	364.6	24.9	645	10	AW691681
33	360	24.6	480	9	AJ476435
34	354.6	24.2	509	13	BJ317630
35	348	23.8	460	10	BE417817
36	347.2	23.7	683	10	AV918966
37	346.8	23.7	454	10	AV915644
38	341.8	23.4	363	10	AW562518
39	335.2	22.9	663	13	BJ241485
40	325.4	22.2	539	12	BF098320
41	317.2	21.7	470	10	AW288831
42	314.8	21.5	376	9	AI065546
43	314.8	21.5	579	10	AV833393
44	314.4	21.5	592	9	AI780966
45	308	21.1	526	13	BI419720

#### ALIGNMENTS

RESULT 1	AY104316	AY104316	1552 bp	linear	HTC 25-MAY-2002
LOCUS	AY104316	Zea mays	PC0066214	mRNA	
DEFINITION	AY104316	Zea mays	PC0066214	mRNA sequence.	
ACCESSION	AY104316				
VERSION	AY104316.1	GI:21207394			
KEYWORDS	HTC.				
SOURCE	HTC.				
ORGANISM	Zea mays.				
	Zea mays				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.				
REFERENCE	1 (bases 1 to 1552)				
AUTHORS	Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whittitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.				
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of.				
JOURNAL	Overgo probes				
REFERENCE	2 (bases 1 to 1552)				
AUTHORS	Coe,E.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
FEATURES	Location/Qualifiers				
source	1. .1552				
	/organism="Zea mays"				
	/db_xref="MaizeDB:633567"				
	/db_xref="taxon:4577"				
	/clone="PC0066214"				

```
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/Note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACS in conjunction with the Maize
Mapping Project"
BASE COUNT 477 a 310 c 381 g 384 t
ORIGIN
Query Match 97.5%; Score 1425.8; DB 11; Length 1552;
Best Local Similarity 99.9%; Pred. No. 8.7e-252;
Matches 1427; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 CGAGAAATAGCTCGCGCGCGGTTTCTTGGCCCACTCCGGCTCAGCCGCCCGCCGCCACC 62
DB 8 CGGAAATAGCTCGCGCGCGGTTTCTTGGCCCACTCCGGCTCAGCCGCCCGCCGCCACC 67
QY 63 CGCCACAGCGCGCGAGAGATGGGATCAAGGGTTTGACGAACTGTCTGGCGGACAA 122
DB 68 CGCCACAGCGCGCGAGAGATGGGATCAAGGGTTTGACGAACTGTCTGGCGGACAA 127
QY 123 TCGCCCAAGCGATGAAGGAGCAGAGTTCGAGAGCTACTTCGCGCGCAAAATCGCGGT 182
DB 128 TCGCGCCAAAGCGATGAAGGAGCAGAGTTCGAGAGCTACTTCGCGCGCAAAATCGCGGT 187
QY 183 CGAGCCCAAGCATATACCAAGTTCCTGATTTGTTGGAAGGAGCAGCATGGAAAC 242
DB 188 CGAGCCCAAGCATATACCAAGTTCCTGATTTGTTGGAAGGAGCAGCATGGAAAC 247
QY 243 TCTCACAATGAAGTGTGAAGTCACTAGTCAATTTGCAAGGAATGTTCAACCGGACAAT 302
DB 248 TCTCACAATGAAGTGTGAAGTCACTAGTCAATTTGCAAGGAATGTTCAACCGGACAAT 307
QY 303 AAGATTACTGGAAGCGGAATCAAGCCAGTTTATGTTTTGTGCGCAAGCTCCTGTATAT 362
DB 308 AAGATTACTGGAAGCGGAATCAAGCCAGTTTATGTTTTGTGCGCAAGCTCCTGTATAT 367
QY 363 GAAGAACAGAGCTTGCTAAAGATACTCAAAAGAGATGATGCAACCAAGATCTGAC 422
DB 368 GAAGAACAGAGCTTGCTAAAGATACTCAAAAGAGATGATGCAACCAAGATCTGAC 427
QY 423 TGAGCGAGTACAGTACAGATGAAGTTCGATGAAAAATTTAGCAAGAGAGCTGTAAA 482
DB 428 TGAGCGAGTACAGTACAGATGAAGTTCGATGAAAAATTTAGCAAGAGAGCTGTAAA 487
QY 483 GGTCAAGGCAACACACAGAGATTTAAACGGCTATTAAAGACTTATGGGGTTCCTGT 542
DB 488 GGTCAAGGCAACACACAGAGATTTAAACGGCTATTAAAGACTTATGGGGTTCCTGT 547
QY 543 TGTAGAGGACCTTCTGAAGCAGAGCAGAAATGTGAGCCCTTTGCAATAACGATAAGGT 602
DB 548 TGTAGAGGACCTTCTGAAGCAGAGCAGAAATGTGAGCCCTTTGCAATAACGATAAGGT 607
QY 603 GTTCGCTGTTGCTTCAGAAATGAGTCCCTACTTTTGGGGCTCCACGGTTCCTTCG 662
DB 608 GTTCGCTGTTGCTTCAGAAATGAGTCCCTACTTTTGGGGCTCCACGGTTCCTTCG 667
QY 663 TCATTTAATGGATCCAAAGTTCGAAGAAATPACCTGTGATGGAATTTGATGTTGCCAAGT 722
DB 668 TCATTTAATGGATCCAAAGTTCGAAGAAATPACCTGTGATGGAATTTGATGTTGCCAAGT 727
QY 723 TTTGGAGGAGCTTGAAGTACCATGACCATGATTCATTTGATGATCCTGTGTGGATG 782
DB 728 TTTGGAGGAGCTTGAAGTACCATGACCATGATTCATTTGATGATCCTGTGTGGATG 787
QY 783 TGACTATTGTGATAGCATCAAGGATTCGGGGCGCAACAGCTCTCAAACTTATTGCTCA 842
DB 788 TGACTATTGTGATAGCATCAAGGATTCGGGGCGCAACAGCTCTCAAACTTATTGCTCA 847
QY 843 ACATGGTCCATGAAGCATCTTGAGAAATCTTTAATAAGACAGATATCAAAATCTCTGA 902
|||||
```

```
Db 848 ACATGGTCCATGAAGCAATCTTGAGAAATCTTAATAAGACAGATATCAAAATCTCTGA 907
QY 903 GGAATGCCCTTACCAAGAGCTCGAGCTTGTTCAGAGGAGCCCTAATGTACATTTGATAT 962
|||||
Db 908 GGAATGCCCTTACCAAGAGCTCGAGCTTGTTCAGAGGAGCCCTAATGTACATTTGATAT 967
QY 963 TCCTGAGCTAAATGGAGCTGCACCTGATGAGGAGGGTCTCATAAAGTTTCTCTGTAAAGA 1022
|||||
Db 968 TCCTGAGCTAAATGGAGCTGCACCTGATGAGGAGGGTCTCATAAAGTTTCTCTGTAAAGA 1027
QY 1023 TAATGTTTTCACGAAGATCGGGTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAA 1082
|||||
Db 1028 TAATGTTTTCACGAAGATCGGGTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAA 1087
QY 1083 TAAATCGTCGCAAGAGAGACTCGAGTCCCTTTTTCAGGCCAATGCCACACATCAGCACC 1142
|||||
Db 1088 TAAATCGTCGCAAGAGAGACTCGAGTCCCTTTTTCAGGCCAATGCCACACATCAGCACC 1147
QY 1143 GCTAAACCGAAGGAGACTTCGGATAAAAACAAGCAAGGCGAGCTGGCAACAGAAAAACAAA 1202
|||||
Db 1148 GCTAAACCGAAGGAGACTTCGGATAAAAACAAGCAAGGCGAGCTGGCAACAGAAAAACAAA 1207
QY 1203 GCGTGTGGAAGAGAAATAATCTTTGGATGCTTGATGTACAACTACGACTACGAAAGCA 1262
Db 1208 GCGTGTGGAAGAGAAATAATCTTTGGATGCTTGATGTACAACTACGACTACGAAAGCA 1267
QY 1263 GCGTGTGCGTGATCACTTCGCTTAGATTTAACTCCTCTGTTTAACTCAGAGCTTTGG 1322
Db 1268 GCGTGTGCGTGATCACTTCGCTTAGATTTAACTCCTCTGTTTAACTCAGAGCTTTGG 1327
QY 1323 TAAAGTTTGTCTCATGTTTTCAGAGCTGGGTAAGTTAGTTGTTTGAAGAGATTGGTGT 1382
Db 1328 TAAAGTTTGTCTCATGTTTTCAGAGCTGGGTAAGTTAGTTGTTTGAAGAGATTGGTGT 1387
QY 1383 CCAAGTAAACAAACTATCGCTGTTTTTACTCTCTGCTTTTGAAGTA 1431
|||||
Db 1388 CCAAGTAAACAAACTATCGCTGTTTTTACTCTCTGCTTTTGAAGTA 1436
```

```
RESULT 2
BG837708 901 bp mRNA linear EST 25-MAY-2001
LOCUS Zm10_01f08_A Zm10_AAF_C_ECOR_Fusarium_graminearum_corn_silk Zea
DEFINITION mays cDNA clone Zm10_01f08, mRNA sequence.
ACCESSION BG837708
VERSION BG837708.1 GI:14204031
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 901)
AUTHORS Harris,L.J., Balcerzak,M., Allard,S., Saparno,A., Couroux,P., De
Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A, Sprott
,D. and Finkler,N.A.
TITLE Expressed Sequence Tags from Maize Silk Six Hours After Silk
Channel Inoculation with Fusarium graminearum
JOURNAL Unpublished (2001)
COMMENT Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harrislj@em.agr.ca.
Location/Qualifiers
1. 901
/organism="Zea mays"
/cultivar="CO388"
/db_xref="taxon:4577"
/clone="Zm10_01f08"
FEATURES
source
```

```
/clone_lib="zm10_AAFc_ECORC_Fusarium_graminearum_corn_silk"
/tissue_type="Silk"
/dev_stage="4-5 days post-silk emergence"
/notes="Vector: Bluescript SK+/XhoI-EcoRI; Site_1: EcoRI; Site_2: XhoI; Field-grown corn was silk channel-inoculated in the morning (~10 am) with 1 ml of a macroconidial suspension (500,000 spores/ml) of Fusarium graminearum and silk channels were collected and immediately frozen in liquid nitrogen 6 hours later. RNA was extracted from silk tissue between 1 cm below and above the inoculation point in the silk channel. RNA from five silk channels was pooled."

BASE COUNT      267 a   174 c   208 g   249 t       3 others
ORIGIN

Query Match      53.2%; Score 778.8; DB 12; Length 901;
Best Local Similarity 98.3%; Pred. No. 3.8e-133;
Matches 816; Conservative 2; Mismatches 9; Indels 3; Gaps 3;

QY 605 TCGCTGTGCTTCAGAGATATGACCTCCTTACTTTTGGGCTCCACGGTCTCTTCGTC 664
DB 1 TCGCTGTGCTTCAGAGATATGACCTCCTTACTTTTGGGCTCCACGGTCTCTTCGTC 60

QY 665 ATTAATGATCCCAAGTTCACAAGAAATACCTGTGATGGAATTTGATGTGCCAAGTTT 724
DB 61 ATTAATGGAYCAAAGTTCACAAGAAATACCTGTGATGGAATTTGATGTGCCAAGTTT 120

QY 725 TGGAGGAGCTTGACCTACCATGACCATGACCTCATTTGATTTGTGATCCTGTGTGGATGTG 784
DB 121 TGGAGGAGCTTGACCTACCATGACCATGACCTCATTTGATTTGTGATCCTGTGTGGATGTG 180

QY 785 ACTATTCTGATAGCATCAAAAGGTATCGGGGGCAACAGCTCTGAACCTTATTCGTCAAC 844
DB 181 ACTATTCTGATAGCATCAAAAGGTATCGGGGGCAACAGCTCTGAACCTTATTCGTCAAC 240

QY 845 ATGGGTCATAGAAAGCATCTTGGAGAACTCTTAATAAGACAGATATCAAAATTC-TCAG 903
DB 241 ATGGGTCATAGAAAGCATCTTGGAGAACTCTTAATAAGACAGATATCAAAATTC-TCAG 300

QY 904 GACTGGCTTACC-AAGAAGCTGACGCTTGTCAAGAGCCCTAATGTCACATTTGGATAT 962
DB 301 GACTGGCTTACC-AAGAAGCTGACGCTTGTCAAGAGCCCTAATGTCACATTTGGATAT 360

QY 963 TCTTGACCTAAATGGAAGTCTGATGAGGAGGCTCTATAAGTTTCTGTTGTTAAAGA 1022
DB 361 TCTTGACCTAAATGGAAGTCTGATGAGGAGGCTCTATAAGTTTCTGTTGTTAAAGA 420

QY 1023 TAATGGTTTCAACGAAGATCGGGTGAACAAGGCCATAGAGAAGATCAAAATCTGCCAAGAA 1082
DB 421 TAATGGTTTCAATGAAGATCGGGTGAACAAGGCCATAGAGAAGATCAAAATCTGCCAAGAA 480

QY 1083 TAATCTGCGCAGGAGACTCGAGTCTTTTCAAGCCACTGCCACCATCAGCACC 1142
DB 481 TAATCTGCGCAGGAGACTCGAGTCTTTTCAAGCCACTGCCACCATCAGCACC 540

QY 1143 GCTAAAAACGGAAGAGACTTCGGATATAAACAAGCAAGCAGCTCGCAACAAGAAACAAA 1202
DB 541 GCTAAAAACGGAAGAGACTTCGGATATAAACAAGCAAGCAGCTCGCAACAAGAAACAAA 600

QY 1203 GGCTGGTGGAAAGAAATAATCTTGGATGCTTGTATGTACAACTACGACTACGAAAGCA 1262
DB 601 GGCTGGTGGAAAGAAATAATCTTGGATGCTTGTATGTACAACTACGACTACGAAAGCA 660

QY 1263 GCGGTGGGTGATCAGTTCGTTAGATATTTAACTCCGTGTTTAACTACAGACTTTGG 1322
DB 661 GCGGTGGGTGATCAGTTCGTTAGATATTTAACTCCGTGTTTAACTACAGACTTTGG 720

QY 1323 TAAAGTTTGTCTCATGTTTCAAGCTGGGTAAAGTTAGTTGTTGTTGAAGAGATTGGTGA 1382
DB 721 TGAAGTTTGTCCCAAGTTTCAAGCTGGGTAAAGTTAGTTGTTGTTGAAGAGATTGGTGA 780

QY 1383 CCAAGTAACAAACTTATCGCTG-TTTTTTACTTCTTGTCTTCTTGAAGTA 1431
```

```
||||| 781 CCAAGTAACAAACTTATCGCTGTTTTTTTACTTCTTGTCTTGAAGTA 830
||||| 781 CCAAGTAACAAACTTATCGCTGTTTTTTTACTTCTTGTCTTGAAGTA 830

RESULT 3
BU037559
LOCUS      BU037559      638 bp      mRNA      linear      EST 23-AUG-2002
DEFINITION      946140C05.v1 946 - tassell primordium prepared by Schmidt lab Zea
               may's cDNA, mRNA sequence.
ACCESSION      BU037559
VERSION        BU037559.1 GI:22473079
KEYWORDS       EST.
SOURCE         Zea mays.
ORGANISM       Zea mays.
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
               clade; Panicoideae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 638)
AUTHORS        Walbot,V.
TITLE          Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY     University
CONTACT        Walbot V
JOURNAL        Unpublished (1999)
COMMENT        Department of Biological Sciences
               Stanford University
               855 California Ave, Palo Alto, CA 94304, USA
               Tel: 650 723 2227
               Fax: 650 725 8221
               Email: walbot@stanford.edu
               Plate: 946140 row: C column: 05.
               Location/Qualifiers
               1..638
               /organism="Zea mays"
               /cultivar="OH43"
               /db_xref="taxon:4577"
               /clone_lib="946 - tassell primordium prepared by Schmidt
               lab"
               /tissue_type="tassels"
               /dev_stage="just after the transition from vegetative to
               inflorescence development"
               /lab_host="XL0LR"
               /note="Organ: tassels; Vector: HybridZAP; Site_1: EcoRI;
               Site_2: XhoI; George Chuck dissected immature tassels
               between 1mm and 3mm. Sharon Stanfield prepared the cDNA
               library in HybridZAP. Sample insert size range was 350 bp
               to 3 Kb with a 1 Kb average."

BASE COUNT      196 a   141 c   171 g   130 t

ORIGIN

Query Match      42.0%; Score 614.8; DB 14; Length 638;
Best Local Similarity 98.7%; Pred. No. 4.9e-103;
Matches 630; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 2 ACAGATATAGTTCGCGCGCGGTTTCTTGGCCACTCCGCTCAGCGCGCGCGCCAC 61
DB 2 ACAGATATAGTTCGCGCGCGGTTTCTTGGCCACTCCGCTCAGCGCGCGCGCCAC 61

QY 62 CCGCCACAGCGCGCGCAGAGATGGCATCAAGGGTTTCACGAACTGCTGG-GGACA 121
DB 62 CCGCCACAGCGCGCGCAGAGATGGCATC-AGGGTTTCACGAACTGCTGGCGGACA 120

QY 122 ATGCGCCCAAGCGGATGAAGGAGCAGAGATTTCGAGAGCTACTTGGCGCGCPAAATCGCG 181
DB 121 ATGCGCCCAAGCGGATGAAGGAGCAGAGATTTCGAGAGCTACTTGGCGCGCPAAATCGCG 180

QY 182 TCGAGCGCAGCATGAGCATATACAGTTCCTGATGTAGTTGGAGAGACAGCATGGAAA 241
DB 181 TCGAGCGCAGCATGAGCATATACAGTTCCTGATAGTAGTTGGAGAGACAGCATGGAAA 240

QY 242 CTCTCACAAATGAAGCTGGTGAAGTCACTACTACTATTTTCAAGGATGTTCAACCGGACAA 301
DB 241 CTCTCACAAATGAAGCTGGTGAAGTCACTACTACTATTTTCAAGGATGTTCAACCGGACAA 300
```

```

QY 302 TAAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTTGTGGCAAGCCTCCTGATA 361
Db 301 TAAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTTGTGGCAAGCCTCCTGATA 360
QY 362 TGAAGAAAACAAGAGCTTCTCTAAAGATACTCAAAAAGAGATGATGCAACCAAGATCTGA 421
Db 361 TGAAGAAAACAAGAGCTTCTCTAAAGATACTCAAAAAGAGATGATGCAACCAAGATCTGA 420
QY 422 CTGAGGCAGTAGAGTAGGAGATAAAGATGCGATTGAAAAATTTGAGCAAGAGGACTGTAA 481
Db 421 CTGAGGCAGTAGAGTAGGAGATAAAGATGCGATTGAAAAATTTGAGCAAGAGGACTGTAA 480
QY 482 AGGTCACAAGGCACACACAACGAAGATTGTAACGCTATTAGACTTTATGGGGGTTCTCG 541
Db 481 AGGTCACAAGGCACACACAACGAAGATTGTAACGCTATTAGACTTTATGGGGGTTCTCG 540
QY 542 TTGTAGAGGCACCTTCTCAAGCAGAAGCAGAAATGTGCAGCCCTTTGCAATAACGATAAGG 601
Db 541 TTGTAGAGGCACCTTCTCAAGCAGAAGCAGAAATGTGCAGCCCTTTGCAATAACGATAAGG 600
QY 602 TGTTCGCTGTTGCTTCAGAAGATATGGACTCCCTTACT 639
Db 601 TGTTCGCTGTTGCTTCAGAAGATATGGACTCCCTTACT 638

RESULT 4
A1881599 586 bp mRNA linear EST 02-FEB-2000
LOCUS 606068G09.y1 606 - Ear tissue cDNA library from Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION A1881599
VERSION A1881599.1 GI:5566733
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
Clade: Panicoideae; Andropogoneae; Zea.
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606068 row: G column: 09.
Location/Qualifiers
1. 586
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt
lab"
/tissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOR (Stratagene)"
/notes="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
lab"

BASE COUNT 186 a 132 c 159 g 109 t
ORIGIN
Query Match 38.38; Score 560.8; DB 9; Length 586;
Best Local Similarity 98.88; Pred. No. 3.9e-93;
Matches 563; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ACGAAGATAGTCGCGCGCGGTTCTTGGCCACATCCGGCTCAGCGCGCGCCGCCAC 61
||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||

```

```

Db 15 ACGGAATAGTCTCGCGGTGCGGGTTCTTGGCCCACTCCGGCTCAGCGCGCGCCGCCAC 74
QY 62 CCGCCACAGCCCGCCGACGAGATGGCATCAAGGGTTTGTACGAAACTGTGGTGGACA 121
Db 75 CCGCCACAGCCCGCCGACGAGATGGCATCAAGGGTTTGTACGAAACTGTGGTGGACA 134
QY 122 ATGGCCCAAGGCGATGAAGGAGCAGAAGTTTCGAGAGCTACTTCGGCCGCGCAAAATCGCGG 181
Db 135 ATGGCCCAAGGCGATGAAGGAGCAGAAGTTTCGAGAGCTACTTCGGCCGCGCAAAATCGCGG 194
QY 182 TCGAGCCAGCATGAGCATATACCAGTTCCCTGATTGTAGTTGGAGGAGCAGCATGGAAA 241
Db 195 TCGAGCCAGCATGAGCATATACCAGTTCCCTGATTGTAGTTGGAGGAGCAGCATGGAAA 254
QY 242 CTCTCACAATTAAGAGCTGGTGAAGTCACTAGTCAATTTCAAGAGGATGTTCAACCGGACAA 301
Db 255 CTCTCACAATTAAGAGCTGGTGAAGTCACTAGTCAATTTCAAGAGGATGTTCAACCGGACAA 314
QY 302 TAAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTTGTGGCAAGCCTCCTGATA 361
Db 315 TAAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTTTGTGGCAAGCCTCCTGATA 374
QY 362 TGAAGAAAACAAGAGCTTCTCTAAAGATACTCAAAAAGAGATGATGCAACCAAGATCTGA 421
Db 375 TGAAGAAAACAAGAGCTTCTCTAAAGATACTCAAAAAGAGATGATGCAACCAAGATCTGA 434
QY 422 CTGAGGCAGTAGAGTAGGAGATAAAGATGCGATTGAAAAATTTGAGCAAGAGGACTGTAA 481
Db 435 CTGAGGCAGTAGAGTAGGAGATAAAGATGCGATTGAAAAATTTGAGCAAGAGGACTGTAA 494
QY 482 AGGTCACAAGGCACACACAACGAAGATTGTAACGCTATTAGACTTTATGGGGGTTCTCG 541
Db 495 AGGTCACAAGGCACACACAACGAAGATTGTAACGCTATTAGACTTTATGGGGGTTCTCG 554
QY 542 TTGTAGAGGCACCTTCTCAAGCAGAAGCAGAA 573
Db 555 TTGTAGAGGCACCTTCTCAAGCAGAAGCAGAA 586

RESULT 5
BE639422 550 bp mRNA linear FST 30-AUG-2000
LOCUS 946033A02.y2 946 - tassell primordium prepared by Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION BE639422
VERSION BE639422.1 GI:9952839
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
Clade: Panicoideae; Andropogoneae; Zea.
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946033 row: A column: 02.
Location/Qualifiers
1. 550
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassell primordium prepared by Schmidt
lab"
/tissue_type="tassels"

FEATURES
Source
1. 550
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassell primordium prepared by Schmidt
lab"
/tissue_type="tassels"

```

```

/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="xLOLR"
/notes="Organ: tassels; Vector: HybrizAP; Site.1: EcoRI;
Site.2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybrizAP. Sample insert size range was 350 bp
to 3 Kb with a 1 Kb average."
177 a 116 c 149 g 108 t
BASE COUNT
ORIGIN

```

BASE COUNT	177 a	116 c	149 g	108 t
ORIGIN				

Query Match	36.7%	Score 536.4	DB 10	Length 550
Best Local Similarity	99.6%	Pred. No. 1.2e-88		
Matches 548	Conservative	0	Mismatches 1	Indels 1
Qy	46	AGCCGCCGCCGCCGCCACACCGCCACAGCCGCGCAGCAGATGGGCATCAAGGGTTTGACG	105	
Db	1	AGCCGCCGCCGCCGCCACACCGCCACAGCCGCGCAGCAGATGGGCATCAAGGGTTTGACG	60	
Qy	106	AAATGCTGGCGGACAAATGCGCCCAAGCGGATGAAGGAGCAGAAAGTTGCGAGAGTACTTC	165	
Db	61	AAATGCTGGCGGACAAATGCGCCCAAGCGGATGAAGGAGCAGAAAGTTGCGAGAGTACTTC	120	
Qy	166	GGCCGCAAAATCGCGTCGCGCCAGCCAGCATGAGCATATACCAGTTCCTCATGTAGTTGGA	225	
Db	121	GGCCGCAAAATCGCGTCGCGCCAGCCAGCATGAGCATATACCAGTTCCTCATGTAGTTGGA	180	
Qy	226	AGGACAGG-CATGGAAACTCTCACAAATGAAGCTGGTGAAGTCACTAGTCAATTTGCAAGG	284	
Db	181	AGGACTGGTCATGGAAACTCTCACAAATGAAGCTGGTGAAGTCACTAGTCAATTTGCAAGG	240	
Qy	285	AATGTTCAACCGGACATAAGATTACTTGGAGCGGGGAATCAAGCCAGTTTATGTTTTGA	344	
Db	241	AATGTTCAACCGGACATAAGATTACTTGGAGCGGGGAATCAAGCCAGTTTATGTTTTGA	300	
Qy	345	TGGCAAGGCTCCTGATATGAAGAAACAGAGCTTGCTAAAAAGATACTCAAAAAGAGATGA	404	
Db	301	TGGCAAGGCTCCTGATATGAAGAAACAGAGCTTGCTAAAAAGATACTCAAAAAGAGATGA	360	
Qy	405	TGCACCAAGATCTGACTGAGGCAGTAGAGGTAGGAGATAAAGATGCGATTGAAAAATT	464	
Db	361	TGCACCAAGATCTGACTGAGGCAGTAGAGGTAGGAGATAAAGATGCGATTGAAAAATT	420	
Qy	465	GAGCAAGGAGGACTGTPAAGGTTCAACGCCACACACCAAGATTCTAAACGCCATTAAAG	524	
Db	421	GAGCAAGGAGGACTGTPAAGGTTCAACGCCACACACCAAGATTCTAAACGCCATTAAAG	480	
Qy	525	ACTTATGGGGTTCCTGTTGTAGGGCCACCTTCTGAAGCAGAAGCAGAAATGTGCAGCCCT	584	
Db	481	ACTTATGGGGTTCCTGTTGTAGGGCCACCTTCTGAAGCAGAAGCAGAAATGTGCAGCCCT	540	
Qy	585	TTGCATAAAC	594	
Db	541	TTGCATAAAC	550	

RESULT 6	
BE639421/c	
LOCUS	554 bp mRNA
DEFINITION	946033A02.x2 946 - tassell primordium prepared by Schmidt lab Zea mays cDNA, mRNA sequence.
	linear EST 30-AUG-2000

ACCESSION	BE639421
VERSION	BE639421.1
	GI:9952838

**KEYWORDS** EST.

**SOURCE** Zea mays.

ORGANISM	Zea mays
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	
11.	
12.	
13.	
14.	
15.	
16.	
17.	
18.	
19.	
20.	
21.	
22.	
23.	
24.	
25.	
26.	
27.	
28.	
29.	
30.	
31.	
32.	
33.	
34.	
35.	
36.	
37.	
38.	
39.	
40.	
41.	
42.	
43.	
44.	
45.	
46.	
47.	
48.	
49.	
50.	
51.	
52.	
53.	
54.	
55.	
56.	
57.	
58.	
59.	
60.	
61.	
62.	
63.	
64.	
65.	
66.	
67.	
68.	
69.	
70.	
71.	
72.	
73.	
74.	
75.	
76.	
77.	
78.	
79.	
80.	
81.	
82.	
83.	
84.	
85.	
86.	
87.	
88.	
89.	
90.	
91.	
92.	
93.	
94.	
95.	
96.	
97.	
98.	
99.	
100.	

REFERENCE  
1 (bases 1 to 554)  
clade; panicoidae; Andropogoneae; Zea.  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; tracheophyta  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; tracheophyta; PACC

**AUTHORS** Walbot, V.

TITLE	Maize EST:
1. <u>Maize</u>	1. <u>Maize</u>
2. <u>Maize</u>	2. <u>Maize</u>
3. <u>Maize</u>	3. <u>Maize</u>
4. <u>Maize</u>	4. <u>Maize</u>
5. <u>Maize</u>	5. <u>Maize</u>
6. <u>Maize</u>	6. <u>Maize</u>
7. <u>Maize</u>	7. <u>Maize</u>
8. <u>Maize</u>	8. <u>Maize</u>
9. <u>Maize</u>	9. <u>Maize</u>
10. <u>Maize</u>	10. <u>Maize</u>
11. <u>Maize</u>	11. <u>Maize</u>
12. <u>Maize</u>	12. <u>Maize</u>
13. <u>Maize</u>	13. <u>Maize</u>
14. <u>Maize</u>	14. <u>Maize</u>
15. <u>Maize</u>	15. <u>Maize</u>
16. <u>Maize</u>	16. <u>Maize</u>
17. <u>Maize</u>	17. <u>Maize</u>
18. <u>Maize</u>	18. <u>Maize</u>
19. <u>Maize</u>	19. <u>Maize</u>
20. <u>Maize</u>	20. <u>Maize</u>
21. <u>Maize</u>	21. <u>Maize</u>
22. <u>Maize</u>	22. <u>Maize</u>
23. <u>Maize</u>	23. <u>Maize</u>
24. <u>Maize</u>	24. <u>Maize</u>
25. <u>Maize</u>	25. <u>Maize</u>
26. <u>Maize</u>	26. <u>Maize</u>
27. <u>Maize</u>	27. <u>Maize</u>
28. <u>Maize</u>	28. <u>Maize</u>
29. <u>Maize</u>	29. <u>Maize</u>
30. <u>Maize</u>	30. <u>Maize</u>
31. <u>Maize</u>	31. <u>Maize</u>
32. <u>Maize</u>	32. <u>Maize</u>
33. <u>Maize</u>	33. <u>Maize</u>
34. <u>Maize</u>	34. <u>Maize</u>
35. <u>Maize</u>	35. <u>Maize</u>
36. <u>Maize</u>	36. <u>Maize</u>
37. <u>Maize</u>	37. <u>Maize</u>
38. <u>Maize</u>	38. <u>Maize</u>
39. <u>Maize</u>	39. <u>Maize</u>
40. <u>Maize</u>	40. <u>Maize</u>
41. <u>Maize</u>	41. <u>Maize</u>
42. <u>Maize</u>	42. <u>Maize</u>
43. <u>Maize</u>	43. <u>Maize</u>
44. <u>Maize</u>	44. <u>Maize</u>
45. <u>Maize</u>	45. <u>Maize</u>
46. <u>Maize</u>	46. <u>Maize</u>
47. <u>Maize</u>	47. <u>Maize</u>
48. <u>Maize</u>	48. <u>Maize</u>
49. <u>Maize</u>	49. <u>Maize</u>
50. <u>Maize</u>	50. <u>Maize</u>
51. <u>Maize</u>	51. <u>Maize</u>
52. <u>Maize</u>	52. <u>Maize</u>
53. <u>Maize</u>	53. <u>Maize</u>
54. <u>Maize</u>	54. <u>Maize</u>
55. <u>Maize</u>	55. <u>Maize</u>
56. <u>Maize</u>	56. <u>Maize</u>
57. <u>Maize</u>	57. <u>Maize</u>
58. <u>Maize</u>	58. <u>Maize</u>
59. <u>Maize</u>	59. <u>Maize</u>
60. <u>Maize</u>	60. <u>Maize</u>
61. <u>Maize</u>	61. <u>Maize</u>
62. <u>Maize</u>	62. <u>Maize</u>
63. <u>Maize</u>	63. <u>Maize</u>
64. <u>Maize</u>	64. <u>Maize</u>
65. <u>Maize</u>	65. <u>Maize</u>
66. <u>Maize</u>	66. <u>Maize</u>
67. <u>Maize</u>	67. <u>Maize</u>
68. <u>Maize</u>	68. <u>Maize</u>
69. <u>Maize</u>	69. <u>Maize</u>
70. <u>Maize</u>	70. <u>Maize</u>
71. <u>Maize</u>	71. <u>Maize</u>
72. <u>Maize</u>	72. <u>Maize</u>
73. <u>Maize</u>	73. <u>Maize</u>
74. <u>Maize</u>	74. <u>Maize</u>
75. <u>Maize</u>	75. <u>Maize</u>
76. <u>Maize</u>	76. <u>Maize</u>
77. <u>Maize</u>	77. <u>Maize</u>
78. <u>Maize</u>	78. <u>Maize</u>
79. <u>Maize</u>	79. <u>Maize</u>
80. <u>Maize</u>	80. <u>Maize</u>
81. <u>Maize</u>	81. <u>Maize</u>
82. <u>Maize</u>	82. <u>Maize</u>
83. <u>Maize</u>	83. <u>Maize</u>
84. <u>Maize</u>	84. <u>Maize</u>
85. <u>Maize</u>	85. <u>Maize</u>
86. <u>Maize</u>	86. <u>Maize</u>
87. <u>Maize</u>	87. <u>Maize</u>
88. <u>Maize</u>	88. <u>Maize</u>
89. <u>Maize</u>	89. <u>Maize</u>
90. <u>Maize</u>	90. <u>Maize</u>
91. <u>Maize</u>	91. <u>Maize</u>
92. <u>Maize</u>	92. <u>Maize</u>
93. <u>Maize</u>	93. <u>Maize</u>
94. <u>Maize</u>	94. <u>Maize</u>
95. <u>Maize</u>	95. <u>Maize</u>
96. <u>Maize</u>	96. <u>Maize</u>
97. <u>Maize</u>	97. <u>Maize</u>
98. <u>Maize</u>	98. <u>Maize</u>
99. <u>Maize</u>	99. <u>Maize</u>
100. <u>Maize</u>	100. <u>Maize</u>

University

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Walbot V  
Department of Biol

Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221

Email: walbot@stanford.edu  
Plate: 946033 row: A column: 02.

FEATURES

LocalIO  
1. .554

/organism="Zea mays"

/cultivar="OH43"

/db\_xref="taxon:4577"

```
/clone_lib="946 - tas
```

```
Lab"
/t+igauo +vno="n+eeo] a
```

```
/dev stage="just after  
/issue_type= laser
```

in/actv\_inflorescence develop

```
/lab_host="XLRLR"
```

/note="Organ: tassels

Site\_2: xhoI; George

between 1mm and 3mm.

RESULT 7  
AW562789

LOCUS

DEFINITION

mRNA sequence.	
ACCESSION	AW562789
VERSION	AW562789.1
KEYWORDS	GI:7216667
SOURCE	EST.
ORGANISM	Zea mays.
	Zea mays.
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 553)
AUTHORS	Walbot.V.
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL	Unpublished (1999)
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 660065 row: H column: 06.
FEATURES	Location/Qualifiers
source	1..553
	/organism="Zea mays"
	/cultivar="Ohio43"
	/db_xref="taxon:4577"
	/clone_lib="660 - Mixed stages of anther and pollen"
	/tissue_type="whole premeiotic anthers to pollen shed"
	/dev_stages="premeiotic anthers to pollen shed"
	/lab_host="XJOLR"
	/note="Organ: Anthers; Vector: Lambda Zap; Site_1: EcoRI; Site_2: XhoI; Anther and pollen cDNA library. Directionally sequenced with 5' end at the EcoRI site. Created by Amie Franklin."
BASE COUNT	171 a 111 c 129 g 142 t
ORIGIN	
Query Match	35.4%; Score 517.8; DB 10; Length 553;
Best Local Similarity	98.7%; Pred. No. 3e-85;
Matches 522; Conservative	0; Mismatches 7; Indels 0; Gaps 0
QY 903	GGACTGGCCCTTACCAAGAGCTCGACGCTTCTTCAAGAGCGCTTAATGTCACATTTGGATAT 962
Db	
4	GGACTGGCCCTTACCAAGAGCTCGACGCTTCTTCAAGAGCGCTTAATGTCACATTTGGATAT 63
QY 963	TCTGAGCTTAAATGGACTGCACTGATGAGGAGGCTCTCATAGTTTCTCGTAAAGA 1022
Db	
64	TCTGAGCTTAAATGGACTGCACTGATGAGGAGGCTCTCATAGTTTCTCGTAAAGA 123
QY 1023	TAAATCGTTCAGGAAGATCGGGTGACAAAGCCATAGAGAATCAAAATCTGCCAAGAA 1082
Db	
124	TAAATCGTTCAGGAAGATCGGGTGACAAAGCCATAGAGAATCAAAATCTGCCAAGAA 183
QY 1083	TAAATCGTTCAGGAAGATCGAGTCTCTTTTCAAGCCAACCTGCCACACATCAGCAC 1142
Db	
184	TAAATCGTTCAGGAAGATCGAGTCTCTTTTCAAGCCAACCTGCCACACATCAGCAC 243
QY 1143	GCTAAACGGAAGAGACTTCGGATAAAACAAGCAGGCGAGCTGCGAACAAAGAAAAACAA 1202
Db	
244	GCTAAACGGAAGAGACTTCGGATAAAACAAGCAGGCGAGCTGCGAACAAAGAAAAACAA 303
QY 1203	GCCTGGTGAAGAAGAAATAATCTTGGATGCTTGATACACTACGACTACGAGAAGCA 1262
Db	
304	GCCTGGTGAAGAAGAAATAATCTTGGATGCTTGATACACTACGACTACGAGAAGCA 363
QY 1263	CGCGTGGCGTGATCACTTCGCTTAGATTAATTTAACTCCCTGTTTAACTCAGAGCTTTGG 1322
Db	
364	CGCGTGGCGTGATCACTTCGCGCTAGATTAATTTAACTCCCTGTTTAACTCAGAGCTTTGG 423
QY 1323	TAAAAGTTTGTCATGTTTCAAGCTGGGGTAAAGTTAGTTGTTTGAAGAGATTGGTGTA 1382
Db	
424	TAAAAGTTTGCCATGTTTCAAGCTGGGGTAAAGTTAGTTGTTTGAAGAGATTGGTGTA 483

Qy	1393	CCAAGTACAAAACTATTCGCTGTGTTTTACTTTCTTCTTGGAAGTA	1431
Db	484	CCAAGTACAAAACTATTCGCTGTGTTTTACTTTCTTCTTGGAAGTA	532
RESULT 8			
AW000375			
LOCUS			
DEFINITION	AW000375	532 bp mRNA linear EST 08-SEP-1999	
		mRNA sequence.	
ACCESSION	AW000375		
VERSION	AW000375.1	GI:5847296	
KEYWORDS	EST.		
SOURCE	Zea mays		
ORGANISM	Zea mays		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC		
	clade; Panicoideae; Andropogoneae; Zea.		
	1 (bases 1 to 532)		
REFERENCE	Walbot, V.		
AUTHORS	Maize ESTs from various cDNA libraries sequenced at Stanford		
TITLE	University		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Walbot V		
	Department of Biological Sciences		
	Stanford University		
	855 California Ave, Palo Alto, CA 94304, USA		
	Tel: 650 723 2227		
	Fax: 650 725 8221		
	Email: walbot@stanford.edu		
	Plate: 614014 row: D column: 03.		
FEATURES	Location/Qualifiers		
source	1..532		
	/organism="Zea mays"		
	/cultivar="W23"		
	/db_xref="taxon:4577"		
	/clone_lib="614 - root cDNA library from Walbot Lab"		
	/tissue_type="root"		
	/dev_stage="3-4 days old"		
	/lab_host="XLOLR"		
	/note="Organ: root; Vector: pBlueScriptII SK+; Site_1:		
	ECORI; Site_2: XhoI; 3-4 days old root tissue from Walbot		
	Lab (LM)"		
BASE COUNT	173 a 103 c 129 g 127 t		
ORIGIN			
	Query Match 34.2%; Score 501; DB 10; Length 532;		
	Best Local Similarity 98.7%; Pred. No. 3.7e-82;		
	Matches 526; Conservative 0; Mismatches 5; Indels 2; Gaps 2;		
QY	666	TTTAATGGATCCAAAGTCCCAAGAAAATACCTGTGATGGAAATTGTATCGGCCAAGGTTTT	725
Db	1	TTTAATGGATCCAAAGTTTCAAGAAAATACCTGTGATGGAAATTGTATCGGCCAAGGTTTT	60
QY	726	GGAGGAGCTTGAACCTCACCATGCACAGTTCATTTGATTTGTCATCCTCTGTGGATGCTGA	785
Db	61	GGAGGAGCTTGAACCTCACCATGCACAGTTCATTTGATTTGTCATCCTCTGTGGATGCTGA	120
QY	786	CTATTGTGATAGCATCAAAGGTATCGGGGGGCAACAGCTCTGAACTTATTCGTCACAAC	845
Db	121	CTATTGTGATAGCATCAAAGGTATCGGGGGGCAACAGCTCTGAACTTATTCGTCACAAC	180
QY	846	TGGGTCCATAGAAGCATCTTGAGAACTCTTAATAAGACA-GAATCAAAATTCCTGAGG	904
Db	181	TGGGTCCATAGAAGCATCTTGAGAACTCTTAATAAGACAGGATATCAAAATTCCTGAAG	240
QY	905	ACTGGCCTTACCAAGAAGCTCGACGCTTTCAAGGAGCCTAATGCACATTCGGATATTC	964
Db	241	ACTGGCCTTACCAAGAAGCTTGAAGCCTTGGTCAAGGAGCCTAATGGC-CATTCGGATATTC	299
QY	965	CTGAGCTAAATGGACTGCACCTGATGAGGAGGCTCTCAATAGTTTCTCTGTTAAAGATA	1024

Db	300	CTGAGCTAAAATGGACTGTCACCTGATGAGGAGGGTCTCATAAGTTTCCTCGTAAAAAGATA	359
Qy	1025	ATGTTTTCACGAAGATCGGGTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAATA	1084
Db	360	ATGGTTTCACGAAGATCGGGTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGAATA	419
Qy	1085	AATCGTGCGAAGGAAGACTCGAGTCTCTTTTTCGAAGCCAACTGCCACACATCATGACGACCGC	1144
Db	420	AATCGTGCGAAGGAAGACTCGAGTCTCTTTTTCGAAGCCAACTGCCACACATCATGACGACCGC	479
Qy	1145	TAAACGGAAGGAGACTTCGGATAAACAAAGCAGGCGAGCTGCGAACCAAGAAA	1197
Db	480	TAAACGGAAGGAGACTTCGGATAAACAAAGCAGGCGAGCTGCGAACCAAGAAA	532
RESULT 9			
AV913663			
LOCUS	AV913663	648 bp mRNA linear	EST 18-JAN-2002
DEFINITION	K. Sato unpublished cDNA library, cv. Haruna Niho germination shoots Hordeum vulgare subsp. vulgare cDNA clone bags22n17 5', mRNA sequence.		
ACCESSION	AV913663		
VERSION	AV913663		
KEYWORDS	EST.		
SOURCE	Hordeum vulgare subsp. vulgare.		
ORGANISM	Hordeum vulgare subsp. vulgare.		
REFERENCE	Eukarya; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; ; Triticeae; Hordeum.		
AUTHORS	Sato, K., Saisho, D. and Takeda, K.		
TITLE	Barley EST sequencing project in NIG and Okayama Univ		
JOURNAL	Unpublished (2002)		
COMMENT	Contact: Tadasu Shin-1 Center for Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.		
FEATURES	location/Qualifiers		
source	1. .648		
	/organism="Hordeum vulgare subsp. vulgare"		
	/cultivar="Haruna Niho"		
	/db_xref="taxon:112509"		
	/clone="bags22n17"		
	/clone_lib="K. Sato unpublished cDNA library, cv. Haruna Niho germination shoots"		
	/tissue_type="shoots"		
	/dev_stage="germination"		
BASE COUNT	174 a	158 c	185 g
ORIGIN			
Query Match	31.8%	Score 465;	DB 10; Length 648;
Best Local Similarity	84.6%;	Pred. No. 1.4e-75;	
Matches 522;	Conservative	0; Mismatches 95;	Indels 0; Gaps 0;
Qy	35	CACATCGGCTCAGCGCGCCGCCACCGCCACAGCGCGCAGACGAGATGGGCATCA	94
Db	32	CTCCCGCCCGCGCGCCTCGCGCGCGCCTCGCGCGCGAGATGGGAGTCA	91
Qy	95	AGGGTTTGACGAAGTCTGTCGGGACAAATGCGCCGAGCGGATGAAGAGCAGAAGTTCG	154
Db	92	AGGGTTTGACGAAGTCTGTCGGGACACACGCGGCCAAGTCGATGAGGAGCAGAAGTTCG	151
Qy	155	AGAGCTACTTCGGCGGCAAAATCGCGCTGACGCGCAGCATGAGATATACAGTTCCCTGA	214
Db	152	AGAGCTACTTCGGCGGCGCATCGCCGTCGACGCCAGCATGAGCATCTACAGTTCCCTTA	211
Qy	215	TGTAGTTTGGAGGACAGCGATGGAACACTCTCACAATAATGAAGCTGGTGAAGTCACTAGTC	274
Db	212	TGTAGTTTGGAGGACAGCGATGGAACACCCCTTACAAACGAAGCGCGGTGATGTCACCAAGTC	271

```
Matches 467; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 920 AAGCTCGAGCGCTGTTCAGGAGCCTAATGTACATTTGATATTCCTGAGCTAAATGGA 979
DB 474 AAGCTCGAGCGCTGTTCAGGAGCCTAATGTACATTTGATATTCCTGAGCTAAATGGA 415
QY 980 CTGCACCTGATGAGGAGGCTCTCATTAAGTTTCTCGTAAAGAGATAATGTTTCAACGAAG 1039
DB 414 CTGCACCTGATGAGGAGGCTCTCATTAAGTTTCTCGTAAAGAGATAATGTTTCAATGAAG 355
QY 1040 ATCGGGTGACAAAGGCCATAGAGAAGATCAATCTGCCAAGATAATATCGTCGCAAGGAA 1099
DB 354 ATCGGGTGACAAAGGCCATAGAGAAGATCAATCTGCCAAGATAATATCGTCGCAAGGAA 295
QY 1100 GACTCGAGTCCTTTTCAAGCCAACTGCCACACATCAGCACCGCTAAAACGGAAGGAGA 1159
DB 294 GACTCGAGTCCTTTTCAAGCCAACTGCCACACATCAGCACCGCTAAAACGGAAGGAGA 235
QY 1160 CTTCGGATAAACAAGCAAGGAGCTGCGAACAAGAAACAAGGCTGGTGGAAAGAGA 1219
DB 234 CTTCGGATAAACAAGCAAGGAGCTGCGAACAAGAAACAAGGCTGGTGGAAAGAGA 175
QY 1220 AATAATCTTGGATGCTTACATCTACAACTACGACTACGAAAGCAGCGTGGCGTGATCACT 1279
DB 174 AATAATCTTGGATGCTTACATCTACAACTACGACTACGAAAGCAGCGTGGCGTGATCACT 115
QY 1280 TCGCTTAGATATTAACTCCCTGTTTAACTCAGAGCTTTGGTAAAGTTTGTCTCATGT 1339
DB 114 TCGCTTAGATATTAACTCCCTGTTTAACTCAGAGCTTTGGTAAAGTTTGTCCATGT 55
QY 1340 TTCAAGCTGGGGTAAAGTTAGTTTGTGTTGAAGAGATTGGTGACCAAGTAAACA 1393
DB 54 TTCAAGCTGGGGTAAAGTTAGTTTGTGTTGAAGAGATTGGTGACCAAGTAAACA 1

RESULT 11
AI861468/c
LOCUS
DEFINITION 614014D03.xl 614 - root cDNA library from Walbot Lab Zea mays cDNA,
AI861468 mRNA sequence.
ACCESSION AI861468
VERSION AI861468.1 GI:5525575
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 470)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL University
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 614014 row: D column: 03.
FEATURES
source
1. .470
Location/Qualifiers
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="614 - root cDNA library from Walbot Lab"
/tissue_type="root"
/dev_stage="3-4 days old"
/lab_host="X10LR"
/note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
Lab (LM)"
```

```
BASE COUNT 112 a 113 c 93 g 152 t
ORIGIN
Query Match 31.6%; Score 462; DB 9; Length 470;
Best Local Similarity 98.9%; Pred. No. 5.4e-75;
Matches 465; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 937 AAGGAGCCTAATGTACATTTGATATTCCTGAGCTAAATGAGCTGCACCTGATGAGAG 996
DB 470 AAGGAGCCTAATGTCCCATTTGATATTCCTGAGCTAAATGAGCTGCCCTGATGAGAG 411
QY 997 GGTCTCATAAAGTTTCTCGTAAAGAGATAATGTTTCAACGAAGATCGGTGACAAAGGCC 1056
DB 410 GGTCTCATAAAGTTTCTCGTAAAGAGATAATGTTTCAACGAAGATCGGTGACAAAGGCC 351
QY 1057 ATAGAGAAGATCAATCTGCCAAGATAATATCGTCGCAAGAGACTCGAGTCCTTTTTC 1116
DB 350 ATAGAGAAGATCAATCTGCCAAGATAATATCGTCGCAAGAGACTCGAGTCCTTTTTC 291
QY 1117 AAGCCAACTGCCACACATCAGCACCGCTAAAACGGAAGGAGCTTCGGATAAAACAAGC 1176
DB 290 AAGCCAACTGCCACACATCAGCACCGCTAAAACGGAAGGAGACTTCGGATAAAACAAGC 231
QY 1177 AAGGAGCTGCGAACAAGAAACAAGGCTGGTGGAAAGAGAAATAATCTTGGATGCTT 1236
DB 230 AAGGAGCTGCGAACAAGAAACAAGGCTGGTGGAAAGAGAAATAATCTTGGATGCTT 171
QY 1237 GATGTACAACTACGACTACGAAAGCAGCGTGGCGTGATCATCTCGCTTAGATTTTAA 1296
DB 170 GATGTACAACTACGACTACGAAAGCAGCGTGGCGTGATCATCTCGCTTAGATTTTAA 111
QY 1297 CTCCTGTTTAACTCAGAGCTTTGGTAAAGTTTGCCTCATGTTTCAAGCTGGGGTAAAGT 1356
DB 110 CTCCTGTTTAACTCAGAGCTTTGGTAAAGTTTGCCTCATGTTTCAAGCTGGGGTAAAGT 51
QY 1357 TAGTTGTTTGAAGAGATTGGTGACCAAGTAAACAATACTATCGCTGT 1406
DB 50 TAGTTGTTTGAAGAGATTGGTGACCAAGTAAACAATACTATCGCTGT 1

RESULT 12
BQ986894
LOCUS
DEFINITION BQ986894
AI861468
ACCESSION BQ986894
VERSION BQ986894.1 GI:22404419
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE 1 (bases 1 to 735)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevallier,P., Ziegler,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
TITLE Lactuce and Sunflower ESTs from the Compositae Genome Project
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozike@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
singleton, see http://cgpdb.ucdavis.edu/ for details.
Plate: QGF10 row: L column: 19.
FEATURES
source
1. .735
Location/Qualifiers
```



```
/organism="Lactuca sativa"  
/cultivar="L. serriola"  
/db_xref="taxon:4236"  
/clone="QGFI0L19"  
/lab_host="E.coli"  
/note="Vector: pBRCDNA5f1AB; The library was constructed  
from 10 different sources of RNA from a single genotype.  
Separate cDNAs were generated using primers that  
incorporated unique 5' and 3' tags to distinguish each  
source of RNA. cDNAs were then pooled, size-fractionated,  
directionally cloned into a custom medium-copy vector and  
transformations made with four size classes to minimize  
size bias. Details of each source of RNA and library  
construction can be obtained at http://cgdb.ucdavis.edu/  
TAG_LIB=QG_EFGHJ lettuce serriola  
TAG_TISSUE=flowers post-fertilized  
TAG_SEQ=GCCATCGGG"  
BASE COUNT 227 a 121 c 178 g 207 t 2 others  
ORIGIN  
Query Match 31.2%; Score 456.4; DB 14; Length 735;  
Best Local Similarity 77.7%; Pred. No. 4.9e-74;  
Matches 550; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
QY 232 GGCATGGAACTCTCACAATGAAGCTGTGAAGTCACTAGTATTTGCAAGGAATGTTTC 291  
Db 1 GGGACAGAAATGCTGACCAATGAGGTGGTGTGCTGACTAGCCATTTGCAAGGAATGTTTC 60  
QY 292 AACCGGACATAGATTACTGGAAGCGGGGAATCAAGCCAGTTTATGTTTGTGATGGCAAG 351  
Db 61 AATCGGACTATTAGGCTTTCTGGAATCTGGATTGAAGCCAGTCTATGTTTGTGATGGGCA 120  
QY 352 CCTCCTGATATGAAGAAACAAGAGCTGTCTAAAGATACCTCAAAAAGAGATGCAACC 411  
Db 121 CCTCCGATTGGAAGAAACAAGACTTGCAAAAGATACCTCAAGCGGAGAGATGCAACA 180  
QY 412 AAGATCTGACTGAGGCGAGTAGAGTAGAGATTAAGATCGGATGGAATAATTCAGCAAG 471  
Db 181 GCGGAGTTAGATGAGGCTATTAAAGTCTGCAATAAGGAACACATGAGAAATTTAGTAA 240  
QY 472 AGGACTGTAAGCTCACAAGGCAACACAGAGATTTGAAGCGCTATTAAAGCTTTATG 531  
Db 241 CGTATGTTGAAGTAAACAAGCAACAATGAGGATTTGCAAAAACCTCTTAAGGCTTATG 300  
QY 532 GGGGTTCTCTGTAGAGGCACCTTCTGAAGCAGAGCAAGATGTGAGCCCTTTGCATA 591  
Db 301 GGAGTGCCTGTTATTGAGGCCCATCTGAAGCAGAGGCACAATGTGCTGCTCTTTGCAAA 360  
QY 592 AACGTAAGGTGTTGCTGTGCTCAGAGATATGGACTCCCTTACTTTTGGGCTCCA 651  
Db 361 GCTGATAAGGTGTATGCTGTGCTCTGAAGATATGGATTCTCTACTTTTGGAGCACCA 420  
QY 652 CGGTTCTTCTGTCATTTAATGATCCCAAGTTCCAGAAATACTCTGTATGGGAATTTGAT 711  
Db 421 AAATTTCTTAGACATTTAATGGATCTGCTCTCAAGAAAATTCCTGTATGGAATTTGAT 480  
QY 712 GTTGCCAAAGTTTGGAGGAGCTTGAACCTCACCATGGACAGCTTCATTTGTTGTCATC 771  
Db 481 GTTTCCAAAGTTTGAAGGAACTGAACCTTACAATGGAATCAATTTATGATCTGTATC 540  
QY 772 CTGTGTGGATGTGACTATTGTGATAGCATCAAAAGGTATCGGGGGGCAACAGCTCTCAA 831  
Db 541 CTATGTGGATGTGATTATTGTGACGATATTAGAGGTATTGTTGGCGACAGACGCTTTGA 600  
QY 832 CTTATTTCGTAACATGGGTCATAGAAAGCATCTTGGAGATCTTTAATAAGACAGATAT 891  
Db 601 CTCATTTCGTAACATGGGTCATAGANACTATCTCTGAAAATATAAACAAGAGAGATAT 660  
QY 892 CAAATTCCTGAGGACTGGGCTTACCAGAGAGCTCGAGCTTTGTTCAAG 939  
Db 661 CANATACCTGAGAAATGGCCATATCAGGAGGCTAGAGCCCTTTTTCAG 708
```

RESULT 13  
BE186786/c

```
LOCUS  
DEFINITION 467 bp mRNA linear EST 22-JUN-2000  
946012C08.X1 946 - tassal primordium prepared by Schmidt lab Zea  
mays cDNA, mRNA sequence.  
ACCESSION BE186786  
VERSION BE186786.1 GI:8665970  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 467)  
Walbot, V.  
Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 946012 row: C column: 08.  
Location/Qualifiers  
1. 467  
/organism="Zea mays"  
/cultivar="OH43"  
/db_xref="taxon:4577"  
/clone_lib="946 - tassal primordium prepared by Schmidt  
lab"  
/tissue_type="tassels"  
/dev_stage="just after the transition from vegetative to  
inflorescence development"  
/lab_host="XL0LR"  
/note="Organ: tassels; Vector: HybriZAP; Site_1: EcoRI;  
Site_2: XhoI; George Chuck dissected immature tassels  
between 1mm and 3mm. Sharon Stanfield prepared the cDNA  
library in HybriZAP. Sample insert size range was 350 bp  
to 3 Kb with a 1 Kb average."  
BASE COUNT 108 a 110 c 96 g 153 t  
ORIGIN  
Query Match 30.6%; Score 448; DB 10; Length 467;  
Best Local Similarity 98.7%; Pred. No. 2e-72;  
Matches 462; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
QY 886 AGATATCAAAATTCCTGAGGACTGGCCCTTACCAGAAGCTGACGCTGTTCAGGAGCCT 945  
Db 467 AGATATCAAAATTCCTGAGGACTGGCCCTTACCAGAAGCTGACGCTGTTCAGGAGCCT 408  
QY 946 AATGTCAATTTGGATATTCCTGAGCTAAATGGACTGACCTGTATGAGGAGGTTCTCAT 1005  
Db 407 AATGTCAATTTGGATATTCCTGAGCTAAATGGACTGACCTGTATGAGGAGGTTCTCAT 348  
QY 1006 AGTTTCTGTGTAAGAATATGTTTCAAGAGATCGGTGACAAAGGCCATAGAGAAG 1065  
Db 347 AGTTTCTGTGTAAGAATATGTTTCAATGAAGTCGGGTGACAAAGGCCATAGAGAAG 288  
QY 1066 ATCAATATCTCCCAAGATAATCGTCAAGGAGACTCGAGTCTCTTTTCAAGCCAACT 1125  
Db 287 ATCAATATCTCCCAAGATAATCGTCAAGGAGACTCGAGTCTCTTTTCAAGCCAACT 228  
QY 1126 GCCACCATATCAGCAGCGCTAAACGGAAGAGACTTCGGATATAAACAAGCAGGAGCT 1185  
Db 227 GCCACCATATCAGCAGCGCTAAACGGAAGAGACTTCGGATATAAACAAGCAGGAGCT 168  
QY 1186 GCCAACAAGAAAACAAGAGCTGGTGGAAAGAAATAATCTTGGATGCTTGTATGACAA 1245  
Db 167 GCCAACAAGAAAACAAGAGCTGGTGGAAAGAAATAATCTTGGATGCTTGTATGACAA 108
```

```

QY 1246 CTAGACTACGAAGCAGCGGTGGCGTGATCAGTTCCTTAGATATATTTAACTCCCGTT 1305
|||||
Db 107 CTAGACTACGAAGCAGCGGTGGCGATGATCAGTTCGC-TAGATATATTTAACTCCCGTT 49
|||||
QY 1306 TTAAGTCTGAGAGCTTTGTGAAAGTTTCTCATGTTTCAAGCTGGGTA 1353
|||||
Db 48 TTAAGTCTGAGAGCTTTGTGAAAGTTTCTCATGTTTCAAGCTGGGTA 1
|||||

RESULT 14
AW562517/c 475 bp mRNA linear EST 10-MAR-2000
LOCUS 660065H06.X2 660 - Mixed stages of anther and pollen Zea mays cDNA,
DEFINITION mRNA sequence.
ACCESSION AW562517.1 GI:7216395
VERSION AW562517.1
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
TITLE clade; Panicoideae; Andropogoneae; Zea.
JOURNAL 1 (bases 1 to 475)
COMMENT Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660065 row: H column: 06.
FEATURES
source
1. .475
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XOLR"
/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."
BASE COUNT 124 a 110 c 94 g 147 t
ORIGIN
Query Match 29.4%; Score 429.8; DB 10; Length 475;
Best Local Similarity 98.2%; Pred. No. 4.3e-69;
Matches 445; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 979 ACTGCACCTGATGAGGAGGCTTCATAGTTTCCTGTGTTAAAGATAATATGTTTCAACGAA 1038
|||||
Db 475 ACTGCACCTGATGAGGAGGCTTCATAGTTTCCTGTGTTAAAGATAATATGTTTCAATGAA 416
|||||
QY 1039 GATCGGTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGATAATATCGTCGCAAGGA 1098
|||||
Db 415 GATCGGTGACAAAGGCCATAGAGAAGATCAAAATCTGCCAAGATAATATCGTCGCAAGGA 356
|||||
QY 1099 AGACTCGAGTCTTTTCAAGCCAACTGCCACCATCAGCAGCCGCTAAAACGGAAGGAG 1158
|||||
Db 355 AGACTCGAGTCTTTTCAAGCCAACTGCCACCATCAGCAGCCGCTAAAACGGAAGGAG 296
|||||
QY 1159 ACTTCGGATAAAACAGCAGGAGCTGCGAACAAAGAAACAAAGGCTGTTGGAAGGAG 1218
|||||
Db 295 ACTTCGGATAAAACAGCAGGAGCTGCGAACAAAGAAACAAAGGCTGTTGGAAGGAG 236
|||||
QY 1219 AAATAATCTTGATGCTTGATGTACACTAGACTACGAAGCAGCGGTGGCGTGATCAC 1278
|||||

```

```

Db 235 AAATAATCTTGATGCTTGATGTACAACTAGACTACGAAGCAGCGGTGGCGTGATCAC 176
|||||
QY 1279 TTCGCTTAGATATATTTAACTCCCTGTTTAACTCAGAGCTTTGGTAAAGTTTCTGCTCATG 1338
|||||
Db 175 TTCGCTTAGATATATTTAACTCCCTGTTTAACTCAGAGCTTTGGTAAAGTTTCTGCTCATG 116
|||||
QY 1339 TTCAAGCTGGGTAAGTTAGTTAGTTGTTTGAAGAGATGGTGTACCAAGTAAACAACTT 1398
|||||
Db 115 TTCAAGCTGGGTAAGTTAGTTAGTTGTTTGAAGAGATGGTGTACCAAGT-ACAAGAACTT 57
|||||
QY 1399 ATCGCTGTTTACTTCTTCTGCTTGAAGTA 1431
|||||
Db 56 ATCGCTGTTTACTTCTTCTGCTTGAAGTA 24
|||||

RESULT 15
BM501417 455 bp mRNA linear EST 14-FEB-2002
LOCUS PAC000000000714 Pioneer AF-1 array Zea mays cDNA, mRNA sequence.
DEFINITION BM501417
ACCESSION BM501417
VERSION BM501417.1 GI:18661475
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
TITLE clade; Panicoideae; Andropogoneae; Zea.
JOURNAL 1 (bases 1 to 455)
COMMENT Hunter,B.G., Beatty,M., Singletary,G., Hamaker,B., Larkins,B.A. and
Jung,R.
Maize opaque endosperm mutations create extensive changes in
patterns of gene expression
Unpublished (2002)
Contact: Jung R
Trait and Technology Development, Food and Feed Research
Pioneer Hi-Bred International, Inc.
7300 NW 62nd Ave., P.O. Box 1004, Johnston, IA 50131-1004, USA
Tel: 515 270 5934
Fax: 515 254 2619
Email: rudolf.jung@pioneer.com.
FEATURES
source
1. .455
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="Pioneer AF-1 array"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI"
BASE COUNT 148 a 81 c 114 g 111 t 1 others
ORIGIN
Query Match 29.0%; Score 423.6; DB 13; Length 455;
Best Local Similarity 98.5%; Pred. No. 5.9e-68;
Matches 448; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 291 CAACGGACATAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTGGATGCA 350
|||||
Db 2 CAACGGACATAGATTACTGGAAGCGGGAATCAAGCCAGTTTATGTTTGGATGCA 61
|||||
QY 351 GCCTCTCATATGAAGAAACAAGAGCTTGTCTAAAGATACCTCAAAAAGAGATGATGCAAC 410
|||||
Db 62 GCCTCTCATATGAAGAAACAAGAGCTTGTCTAAAGATACCTCAAAAAGAGATGATGCAAC 121
|||||
QY 411 CAAAGATCTGACTGAGGAGTAGAGGTAGAGATAAGATCGGATTTGAAATTTGAGCA 470
|||||
Db 122 CAAAGATCTGACTGAGGAGTAGAGGTAGAGATAAGATCGGATTTGAAATTTGAGCA 181
|||||
QY 471 GAGGACTCTAAGCTCAAGGCAACCAAGAGATTGTAACGGCTATTAAAGACTTAT 530
|||||
Db 182 GAGGACTCTAAGGTCACAGGCAACCAAGAGATTGTAACGGCTATTAAAGACTTAT 241
|||||
QY 531 GGGGTTCTCTGTTAGAGGCACTTCTTGAGCAGAGAGAGAGATGTCAGCCCTTTGCAT 590
|||||
Db 242 GGGGTTCTCTGTTAGAGGCACTTCTTGAGCAGAGAGAGAGATGTCAGCCCTTTGCAT 301
|||||

```

```

Qy 591 AAACGATAAGGTGTTGCTGTTGCTTCAGAAAGATATGGAATCCCTTACTTTTGGGGCTCC 650
Db 302 AAACGATAAGGTGTTGCTGTTGCTTCAGAAAGATATGGAATCCCTTACTTTTGGGGCTCC 361
Qy 651 ACGGTTCCCTGCTCATTAAAT-GGATCCAAAGTTCCAAAGAAAATACCTGTGATGGAATTTG 709
Db 362 ACGGTTCCCTGCTCATTAAATGGGATCCAAAGTTCCAAAGAAAATACCTGTGATGGAATTTG 421
Qy 710 ATGTTGCCAAGGTTTGGAGGAGCTTGAACCTCAC 744
Db 422 ATGTTGCCAAGG-TTGGAGGAGCCTTAACCTCANC 455

```

Search completed: May 8, 2003, 21:34:20  
 Job time : 1973 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 19:58:42 ; Search time 62 Seconds  
(without alignments)  
814,548 Million cell updates/sec

Title: US-09-805-311-2

Perfect score: 1939

Sequence: 1 MGIKGLTKLADNAPKAMKE.....SDKTSKAAANKTKAGGKKK 379

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1939	100.0	379	21	AAV95307
2	1939	100.0	379	21	AAV95309
3	1933	99.7	379	21	AAV95308
4	1933	99.7	379	21	AAV95310
5	1031	53.2	380	20	AAW92504
6	1007.5	52.0	377	20	AAW92505
7	998.5	51.5	385	22	AB963960
8	996	51.4	378	20	AAW92508
9	938	48.4	382	20	AAW92506
10	699.5	36.1	373	22	ABG19545

11	648.5	33.4	340	18	AAW24216	Pyrococcus furiosu
12	648.5	33.4	340	19	AAW79970	Pyrococcus furiosu
13	648.5	33.4	340	19	AAW59940	Amino acid sequenc
14	647	33.4	326	19	AAW59953	Amino acid sequenc
15	644	33.2	332	19	AAW59951	Amino acid sequenc
16	636	32.8	343	22	AAW96644	Putative P. abyssal
17	632.5	32.6	343	20	AAW03778	Pyrococcus heat re
18	609	31.4	325	19	AAW59950	Amino acid sequenc
19	600	30.9	340	19	AAW59949	Amino acid sequenc
20	562	29.0	340	19	AAW59952	Amino acid sequenc
21	560.5	28.9	326	18	AAW24215	Methanococcus jann
22	560.5	28.9	326	19	AAW59939	Amino acid sequenc
23	553.5	28.5	326	19	AAW79969	Methanococcus jann
24	538.5	27.8	336	19	AAW79982	Archaeoglobus fulg
25	538.5	27.8	336	19	AAW59946	Amino acid sequenc
26	525	27.1	328	19	AAW59948	Amino acid sequenc
27	411	21.2	258	19	AAW59947	yeast delta-RAD2 p
28	340.5	17.6	386	20	AAW92507	Plasmodium faicipa
29	336	17.3	1516	21	AB18195	Human nuclease-lik
30	311	16.0	96	23	ABP33304	Drosophila melanog
31	257	13.3	726	22	AB965192	Drosophila melanog
32	243.5	12.6	1236	22	AB965297	Drosophila melanog
33	243.5	12.6	1257	22	AB967273	Thermophilic DNA p
34	215.5	11.1	871	23	AAW48927	Thermophilic DNA p
35	215.5	11.1	871	23	AAW48928	Drosophila melanog
36	210	10.8	732	22	AB971917	Drosophila melanog
37	208	10.7	872	20	AAV31815	Thermoanaerobacter
38	205	10.6	872	18	AAW26605	Thermoanaerobacter
39	205	10.6	872	20	AAV31812	T. thermohydrosulf
40	205	10.6	872	20	AAV31816	T. caldophilus DNA
41	184.5	9.5	291	20	AAW78466	Taq polymerase enc
42	181	9.3	832	13	AAW22602	Taq polymerase enc
43	174.5	9.0	832	13	AAW22603	Taq polymerase enc
44	174.5	9.0	832	13	AAW22604	Taq polymerase enc
45	173.5	8.9	834	17	AAW99542	Thermus flavus DNA

#### ALIGNMENTS

#### RESULT 1

AAV95307  
ID AAV95307 standard; Protein; 379 AA.

XX AAV95307;

XX AAV95307;

DT 12-SEP-2000 (first entry)

XX Maize Rad2/FEN-1 protein.

DE Maize; Rad2/FEN-1; transgenic plant; male sterile plant;

XX endonuclease; exonuclease; DNA repair; gene targeting.

XX endonuclease; exonuclease; DNA repair; gene targeting.

OS Zea mays.

XX WO200036109-A1.

XX 22-JUN-2000.

XX 16-NOV-1999; 99WO-US27147.

XX 15-DEC-1998; 98US-0112332.

XX (PION-) PIONEER HI-BRED INT INC.

XX Mahajan PB;

XX WPI; 2000-452026/39.

XX N-PSDB; AAA27923.

XX Maize RAD2/FEN-1 nucleic acids and proteins useful for modulating DNA

PT recombination and repair in transgenic plants, e.g. for gene targeting

PT and the production of male sterile plants -

XX Claim 11; Page 71-72; 85pp; English.

XX The present sequence is that of maize Rad2/FEN-1, as deduced from

CC a cDNA clone (see AAA27923) isolated from maize line B73 immature

CC ear tissue. Rad2/FEN-1 is a structure specific endonuclease which

CC under certain conditions also acts as an exonuclease. Rad2/FEN-1

CC can be expressed in transgenic plant cells using conventional

CC methods. The protein is involved in the regulation of DNA repair

CC and recombination in plant systems and therefore may be used for

CC improving gene targeting during further recombinant DNA protocols

CC involving plants. Rad2/FEN-1 endonucleolytic activity is essential

CC in DNA replication and nucleotide excision and repair reactions.

CC The exolytic activity is involved in double strand break repair and

CC end joining. The protein is also useful in strand exchange

CC reactions during homologous recombination. These functions may be

CC useful in gene targeting and in the production of male sterile

CC plants. The efficacy of gene targeting can be improved by the

CC overexpression of exogenous Rad2/FEN-1 while male sterile plants can

CC be produced by the down regulation of Rad2/FEN-1 expression.

XX Sequence 379 AA;

Query Match 100.0%; Score 1939; DB 21; Length 379;

Best Local Similarity 100.0%; Pred. No. 1.4e-170;

Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGIGLTKLLADNAPKAMKEQKFSYFGKTIADVADSMISYQFLIYVGRGTMETLTNEAGE 60

Db 1 MGIGLTKLLADNAPKAMKEQKFSYFGKTIADVADSMISYQFLIYVGRGTMETLTNEAGE 60

Qy 61 VTSHLQGMFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

Db 61 VTSHLQGMFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

Qy 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMLGVPVVEAPSEAEACALCINDKVFVASED 180

Db 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMLGVPVVEAPSEAEACALCINDKVFVASED 180

Qy 181 MDSLTFGAPRFLRHLMDSPPSKKIPVMEFDVAKVLEELTMDQFDLCLCGCDYCDISIK 240

Db 181 MDSLTFGAPRFLRHLMDSPPSKKIPVMEFDVAKVLEELTMDQFDLCLCGCDYCDISIK 240

Qy 241 GIGGTALKLIRHQHSIESILENLNDRYQIPEDWPYQEARRLFKEPNVTLIDPELKWTA 300

Db 241 GIGGTALKLIRHQHSIESILENLNDRYQIPEDWPYQEARRLFKEPNVTLIDPELKWTA 300

Qy 301 PDEGLISFLVKDNGFNEDRVTKAIEKISAKNKSQGRLESFFKPTATTAPLKRKETS 360

Db 301 PDEGLISFLVKDNGFNEDRVTKAIEKISAKNKSQGRLESFFKPTATTAPLKRKETS 360

Qy 361 DKTSKAAANKTKAGGKKK 379

Db 361 DKTSKAAANKTKAGGKKK 379

RESULT 2

AA95309

ID AA95309 standard; Protein; 379 AA.

XX

AC AA95309;

XX

DT 12-SEP-2000 (first entry)

XX

DE Maize Rad2/FEN-1 protein.

XX

KW Maize; Rad2/FEN-1; transgenic plant; male sterile plant;

KW endonuclease; exonuclease; DNA repair; gene targeting.

XX

OS Zea mays.

XX

PN WO200036109-A1.

XX

PD 22-JUN-2000.

XX

PF 16-NOV-1999; 99WO-US27147.

XX

PR 15-DEC-1998; 98US-0112332.

XX

PA (PTON-) PIONEER HI-BRED INT INC.

XX

PI Mahajan PB;

XX

DR WPI; 2000-452026/39.

DR N-FSDB; AAA27925.

XX

Maize Rad2/FEN-1 nucleic acids and proteins useful for modulating DNA

PT recombination and repair in transgenic plants, e.g. for gene targeting

PT and the production of male sterile plants -

XX

PS Example 1; Page 76-77; 85pp; English.

XX

CC The present sequence is that of maize Rad2/FEN-1, as deduced from

CC a cDNA clone (see AAA27925) derived from maize line W23 tassel

CC polyA RNA. Rad2/FEN-1 is a structure specific endonuclease which

CC under certain conditions also acts as an exonuclease. Rad2/FEN-1

CC can be expressed in transgenic plant cells using conventional

CC methods. The protein is involved in the regulation of DNA repair

CC and recombination in plant systems and therefore may be used for

CC improving gene targeting during further recombinant DNA protocols

CC involving plants. Rad2/FEN-1 endonucleolytic activity is essential

CC in DNA replication and nucleotide excision and repair reactions.

CC The exolytic activity is involved in double strand break repair and

CC end joining. The protein is also useful in strand exchange

CC reactions during homologous recombination. These functions may be

CC useful in gene targeting and in the production of male sterile

CC plants. The efficacy of gene targeting can be improved by the

CC overexpression of exogenous Rad2/FEN-1 while male sterile plants can

CC be produced by the down regulation of Rad2/FEN-1 expression.

XX

SQ Sequence 379 AA;

Query Match 100.0%; Score 1939; DB 21; Length 379;

Best Local Similarity 100.0%; Pred. No. 1.4e-170;

Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGIGLTKLLADNAPKAMKEQKFSYFGKTIADVADSMISYQFLIYVGRGTMETLTNEAGE 60

Db 1 MGIGLTKLLADNAPKAMKEQKFSYFGKTIADVADSMISYQFLIYVGRGTMETLTNEAGE 60

Qy 61 VTSHLQGMFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

Db 61 VTSHLQGMFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

Qy 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMLGVPVVEAPSEAEACALCINDKVFVASED 180

Db 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMLGVPVVEAPSEAEACALCINDKVFVASED 180

Qy 181 MDSLTFGAPRFLRHLMDSPPSKKIPVMEFDVAKVLEELTMDQFDLCLCGCDYCDISIK 240

Db 181 MDSLTFGAPRFLRHLMDSPPSKKIPVMEFDVAKVLEELTMDQFDLCLCGCDYCDISIK 240

Qy 241 GIGGTALKLIRHQHSIESILENLNDRYQIPEDWPYQEARRLFKEPNVTLIDPELKWTA 300

Db 241 GIGGTALKLIRHQHSIESILENLNDRYQIPEDWPYQEARRLFKEPNVTLIDPELKWTA 300

Qy 301 PDEGLISFLVKDNGFNEDRVTKAIEKISAKNKSQGRLESFFKPTATTAPLKRKETS 360

Db 301 PDEGLISFLVKDNGFNEDRVTKAIEKISAKNKSQGRLESFFKPTATTAPLKRKETS 360

Qy 361 DKTSKAAANKTKAGGKKK 379

Db 361 DKTSKAAANKTKAGGKKK 379

RESULT 3

```
AAAY95308
ID AAY95308 standard; Protein; 379 AA.
XX
AC AAY95308;
XX
DT 12-SEP-2000 (first entry)
XX
DE Maize Rad2/FEN-1 protein.
XX
KW Maize; Rad2/FEN-1; transgenic plant; male sterile plant;
KW endonuclease; exonuclease; DNA repair; gene targeting.
XX
OS Zea mays.
XX
PN WO200036109-A1.
XX
PD 22-JUN-2000.
XX
PF 16-NOV-1999; 99WO-US27147.
XX
PR 15-DEC-1998; 98US-0112332.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Mahajan PB;
XX
XX WPI; 2000-452026/39.
DR N-PSDB; AAA27924.
XX
XX Maize RAD2/FEN-1 nucleic acids and proteins useful for modulating DNA
PT recombination and repair in transgenic plants, e.g. for gene targeting
PT and the production of male sterile plants -
XX
PS Example 1; Page 74; 85pp; English.
XX
CC The present sequence is that of maize RAD2/FEN-1, as deduced from
CC a cDNA clone (see AAA27924) derived from maize line B73 seedling
CC tissue RNA. Rad2/FEN-1 is a structure specific endonuclease which
CC under certain conditions also acts as an exonuclease. Rad2/FEN-1
CC can be expressed in transgenic plant cells using conventional
CC methods. The protein is involved in the regulation of DNA repair
CC and recombination in plant systems and therefore may be used for
CC improving gene targeting during further recombinant DNA protocols
CC involving plants. RAD2/FEN-1 endonucleolytic activity is essential
CC in DNA replication and nucleotide excision and repair reactions.
CC The exolytic activity is involved in double strand break repair and
CC end joining. The protein is also useful in strand exchange
CC reactions during homologous recombination. These functions may be
CC useful in gene targeting and in the production of male sterile
CC plants. The efficacy of gene targeting can be improved by the
CC overexpression of exogenous Rad2/FEN-1 while male sterile plants can
CC be produced by the down regulation of Rad2/FEN-1 expression.
XX
SQ Sequence 379 AA;

Query Match 99.7%; Score 1933; DB 21; Length 379;
Best Local Similarity 99.7%; Pred. No. 5e-170;
Matches 378; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGIKGUTKLLADNAPKAMKEQFESYFGRIKIAVDASMSIYQFLIVVGRGTGMETLTNEAGE 60
|||||
DB 1 MGIKGLTKLLADNAPKAMKEQFESYFGRIKIAVDASMSIYQFLIVVGRGTGMETLTNEAGE 60
|||||

QY 61 VTSHLQGMENRTTLLEAGIKPYYVDGKPPDMKKQBLAKRYSKRDDATKDLTAEVVGVD 120
|||||
DB 61 VTSHLQGMENRTTLLEAGIKPYYVDGKPPDMKKQBLAKRYSKRDDATKDLTAEVVGVD 120
|||||

QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPPVVEAPSEAEACALCINDKVFVASED 180
|||||
DB 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPPVVEAPSEAEACALCINDKVFVASED 180
|||||

QY 181 MSLTFTGAPRFLRLMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCLCGDCYDSIK 240
|||||
DB 181 MSLTFTGAPRFLRLMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCLCGDCYDSIK 240
|||||

Db 181 MSLTFTGAPRFLRLMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCLCGDCYDSIK 240
QY 241 GIGGOTALKLIROHGSIESILENLNKRQYQIPEDWPYQEARLRFKEPNVTLDIPELKWTA 300
|||||
Db 241 GIGGOTALKLIROHGSIESILENLNKRQYQIPEDWPYQEARLRFKEPNVTLDIPELKWTA 300
|||||
QY 301 PDEEGLISFLVKDNGFNEDRVTKAIEKISAKNKSQGRLESFFFKPTATTSAFLKRRKETS 360
|||||
Db 301 PDEEGLISFLVKDNGFNEDRVTKAIEKISAKNKSQGRLESFFFKPTATTSAFLKRRKETS 360
|||||
QY 361 DKTSKAAANKTKAGGKKK 379
|||||
Db 361 DKTSKAAANKTKAGGKKK 379

RESULT 4
AAY95310
ID AAY95310 standard; Protein; 379 AA.
XX
AC AAY95310;
XX
DT 12-SEP-2000 (first entry)
XX
DE Maize Rad2/FEN-1 protein.
XX
KW Maize; Rad2/FEN-1; transgenic plant; male sterile plant;
KW endonuclease; exonuclease; DNA repair; gene targeting.
XX
OS Zea mays.
XX
PN WO200036109-A1.
XX
PD 22-JUN-2000.
XX
PF 16-NOV-1999; 99WO-US27147.
XX
PR 15-DEC-1998; 98US-0112332.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Mahajan PB;
XX
XX WPI; 2000-452026/39.
DR N-PSDB; AAA27926.
XX
XX Maize RAD2/FEN-1 nucleic acids and proteins useful for modulating DNA
PT recombination and repair in transgenic plants, e.g. for gene targeting
PT and the production of male sterile plants -
XX
PS Example 1; Page 79-80; 85pp; English.
XX
CC The present sequence is that of maize RAD2/FEN-1, as deduced from
CC a cDNA clone (see AAA27926) derived from maize line B73 endosperm
CC RNA. Rad2/FEN-1 is a structure specific endonuclease which
CC under certain conditions also acts as an exonuclease. Rad2/FEN-1
CC can be expressed in transgenic plant cells using conventional
CC methods. The protein is involved in the regulation of DNA repair
CC and recombination in plant systems and therefore may be used for
CC improving gene targeting during further recombinant DNA protocols
CC involving plants. RAD2/FEN-1 endonucleolytic activity is essential
CC in DNA replication and nucleotide excision and repair reactions.
CC The exolytic activity is involved in double strand break repair and
CC end joining. The protein is also useful in strand exchange
CC reactions during homologous recombination. These functions may be
CC useful in gene targeting and in the production of male sterile
CC plants. The efficacy of gene targeting can be improved by the
CC overexpression of exogenous Rad2/FEN-1 while male sterile plants can
CC be produced by the down regulation of Rad2/FEN-1 expression.
XX
SQ Sequence 379 AA;

Query Match 99.7%; Score 1933; DB 21; Length 379;
Best Local Similarity 99.7%; Pred. No. 5e-170;
Matches 378; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```





PT DNA encoding flap endonuclease polypeptides - useful for producing  
PT e.g. recombinant polypeptides  
PS Claim 1; Fig 2A; 58pp; English.  
XX This sequence represents a mouse FEN-1 (flap endonuclease) protein. This  
CC protein can be used in methods for detecting a pathological condition in  
CC a patient, for diagnostic purposes, for screening for antineoplastic  
CC agents and carcinogens, for diagnostic staging of neoplasia, for  
CC producing recombinant flap endonuclease for use as research or  
CC diagnostic reagents, for producing antibodies reactive with the novel  
CC polypeptides, for producing transgenic nonhuman animals expressing the  
CC novel polypeptides encoded by a transgene. The invention also provides  
CC novel molecular cloning techniques and reagents involving cleavage of  
CC a flap or nick with a flap endonuclease.  
XX  
SQ Sequence 377 AA;  
Query Match 52.4%; Score 1007.5; DB 20; Length 377;  
Best Local Similarity 53.4%; Pred. No. 1.7e-84;  
Matches 203; Conservative 68; Mismatches 100; Indels 9; Gaps 7;  
QY 1 MGKGLTKLLADNAPKAMKQKESYFGRKIAVDASMSIYQFLIVGRTGMETLTNEAGE 60  
Db 1 MGTHGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIV-RQGGDVQLQNEGE 59  
QY 61 VTSHLQGMFNRTIRLEAGIKPVYVFDGKPPDKMKQELAKRYSKRDDATKDLTEAVEVD 120  
Db 60 TTS-LMGMFYRTIR-MENGIKPVYVFDGKPPQLKSGELAKRSERRAEKLOQAQAG- 116  
QY 121 KDAIEKLSKRVKVTROHNECKRLLRLMGVVPVVEAPSEAECAALCINDKVPFAVASED 180  
Db 117 MEEVEKFKRLVKVTKOHNECKRLLSLMGIPYLDAPSEAECAALAKAGKYVAATED 176  
QY 181 MDSITFGAPRFLRHLMDFSSKKIPVMEFDVAKVLEELTMDQFIDLCILGCDYCSIK 240  
Db 177 MDCLTFGSPVLMRHLTASEAKKLPQEFHLSRVQLQELGLAQEQFVDCILIGSDYCESIR 236  
QY 241 GIGGQTKLIRHSGSTESILENLNKDRYQIPEDMPYQEARLKFEPNVT-LDIPELKWT 299  
Db 237 GIGAKRAVDLIQKHKSIEEIVRRLDPSKYPVFNWHLKHAQQLFEPEVDPESELKWS 296  
QY 300 APDEGLISFLVKDNGFNEEDRVTKAIEKISAKNKSQGRLESFFKPTATTSAPLKRKET 359  
Db 297 EPNEELVKFCGKQFSEERISGVKRLSKRSRGSTQGRLLDFKFTGSLSS-AKRKEP 355  
QY 360 SDKTSKAAANKTKTAGKKK 379  
Db 356 EPKGS---AKKKARTGGAGK 372  
RESULT 7  
AB63960  
ID ABB63960 standard; Protein; 385 AA.  
XX  
AC ABB63960;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 18672.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX  
PN 27-SEP-2001.  
XX  
PD 23-MAR-2001; 2001WO-US09231.  
XX  
PF 23-MAR-2000; 2000US-191637P.  
XX  
PR

PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
PA Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
PI N-PSDB; ABL08063.  
XX  
DR  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 18672; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 385 AA;  
Query Match 51.5%; Score 998.5; DB 22; Length 385;  
Best Local Similarity 52.9%; Pred. No. 1.2e-83;  
Matches 202; Conservative 65; Mismatches 104; Indels 11; Gaps 5;  
QY 1 MGKGLTKLLADNAPKAMKQKESYFGRKIAVDASMSIYQFLIVGRTGMETLTNEAGE 60  
Db 1 MGTHGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIVRSEGAQ-LATVNGD 59  
QY 61 VTSHLQGMFNRTIRLEAGIKPVYVFDGKPPDKMKQELAKRYSKRDDATKDLTEAVEVD 120  
Db 60 PTHLMGMFRTIRLLDNGIKPVYVFDGKPPDLKSGELAKRAERREAEKLLKAATDAGD 119  
QY 121 KDAIEKLSKRVKVTROHNECKRLLRLMGVVPVVEAPSEAECAALCINDKVPFAVASED 180  
Db 120 DAGIEKFNRLRVVTKHAKAEKELLTLMGVYVVDAPCEAEQAQCAALYKAGKVATATED 179  
QY 181 MDSITFGAPRFLRHLMDFSSKKIPVMEFDVAKVLEELTMDQFIDLCILGCDYCSIK 240  
Db 180 MDALTFGSTKRLRYLTSEARKMPVKESYDKLLEGLAINNREFIDLCILGCDYCESIK 239  
QY 241 GIGGQTKLIRHSGSTESILENLNKDRYQIPEDMPYQEARLKFEPNVT-LDIPELKWT 299  
Db 240 GIGPKRAIELINTYRDIETILDNLDSKSYTPENNWKVARELFTPEVADADSDLKWV 299  
QY 300 APDEGLISFLVKDNGFNEEDRVTKAIEKISAKNKSQGRLESFFK----PTATTSAPL 354  
Db 300 EPDEGLVFKLCGRQFNEERVNAGKMLKMSKQATQVRLLDSFFKTLFSTPNATAA-- 357  
QY 355 KRKETSDDTKSAAANKTKTAGK 376  
Db 358 KRK--AEAKKSANNKAKTSG 377  
RESULT 8  
AAW92508  
ID AAW92508 standard; Protein; 378 AA.  
XX  
AC AAW92508;  
XX  
DT 23-APR-1999 (first entry)  
XX  
DE Human FEN-1 protein.  
XX  
KW FEN-1; human; flap endonuclease; detection; diagnosis; carcinogen;

KW neoplasma; antineoplastic agent; cleavage.

OS Homo sapiens.

XX US5874283-A.

XX 23-FEB-1999.

XX 30-MAY-1995; 95US-0455968.

XX 30-MAY-1995; 95US-0455968.

XX (HARR/) HARRINGTON J J.

XX (HSIEH/) HSIEH C.

XX (LIEB/) LIEBER M R.

XX Harrington JJ, Hsieh C, Lieber MR;

XX WPI: 1999-179985/15.

XX N-PSDB; AAX02111.

XX DNA encoding flap endonuclease polypeptides - useful for producing

XX e.g. recombinant polypeptides

XX Disclosure; Fig 5A-B; 58pp; English.

XX This sequence represents a human FEN-1 (flap endonuclease) protein. This protein can be used in methods for detecting a pathological condition in a patient, for diagnostic purposes, for screening for antineoplastic agents and carcinogens, for diagnostic staging of neoplasia, for producing recombinant flap endonuclease for use as research or diagnostic reagents, for producing antibodies reactive with the novel polypeptides, for producing transgenic nonhuman animals expressing the novel polypeptides encoded by a transgene. The invention also provides novel molecular cloning techniques and reagents involving cleavage of a flap or nick with a flap endonuclease.

XX Sequence 378 AA;

Query Match 51.4%; Score 996; DB 20; Length 378;

Best Local Similarity 52.4%; Pred. No. 2e-83;

Matches 199; Conservative 70; Mismatches 103; Indels 8; Gaps 6;

QY 1 MGIGLTKLLADNAPKAMKEQKFESEYFGKRTAVDASMSIYQFLIVVGRGTMETLTNEAGE 60

Db 1 MEIHLAKLIADVAPSAIRENDIKSYFGKRTAVDASMSIYQFLIAY-RQGGDVQLQNEEGE 59

QY 61 VTSLQGMFNRTIRLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

Db 60 TTS-LMGMFYRTMR-MENGIKPVYVFDGKPPOLKSGELAKRSERRAEKQLQQAQAGM 117

QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPPVDPAPSEAEACALCINDKVFVASED 180

Db 118 EEEVEKFTKRLVKVTKQHNEDCKHLLMGIPYLDLPAPSEAEACALAKAGKVVAAATED 177

QY 181 MDSLTFGAPRFLRHLMOPSSKKIPWMEFDVAKVLEELTMDQFDLCLCGCDYCDISK 240

Db 178 MDCLTFGSPVLMRLHMTASEAKLPQIEFHLKVLQELGNGEQFVLDCLIGSDYCESIR 237

QY 241 GIGGOTALKLIROHGSIESILENLKDRYQIPEDWPYQEARLFKEPNVT-LDIPELKWT 299

Db 238 GIGAKRAVDLIQHKHSIEEIVRLDPSKYPPVFNWHLKEAQQLFLEPVDPESVELAWS 297

QY 300 APDEGLISFLVKDNGFNEDRVTKAIEKISKAKNKSOGRLSEFPKPTATTSAPLKRKET 359

Db 298 EPNEEELVKFCGKEQFFFEIRSGVKRLSKSRQSGTQGRLDFFKVTGSLSS-AKRKEP 356

QY 360 SDKTSKAAANKTKAGGKKK 379

Db 357 E---PKGPAKKKATGGAGK 373

RESULT 9

AAW92506

ID AAW92506 standard; Protein; 382 AA.

XX AAW92506;

XX 23-APR-1999 (first entry)

XX Yeast FEN-1 protein.

XX FEN-1; yeast; flap endonuclease; detection; diagnosis; carcinogen;

XX neoplasma; antineoplastic agent; cleavage.

XX Saccharomyces cerevisiae.

XX US5874283-A.

XX 23-FEB-1999.

XX 30-MAY-1995; 95US-0455968.

XX 30-MAY-1995; 95US-0455968.

XX (HARR/) HARRINGTON J J.

XX (HSIEH/) HSIEH C.

XX (LIEB/) LIEBER M R.

XX Harrington JJ, Hsieh C, Lieber MR;

XX WPI: 1999-179985/15.

XX N-PSDB; AAX02109.

XX DNA encoding flap endonuclease polypeptides - useful for producing

XX e.g. recombinant polypeptides

XX Disclosure; Fig 3A; 58pp; English.

XX This sequence represents a yeast FEN-1 (flap endonuclease) protein. This protein is used in a method to isolate novel human FEN-1 proteins for detecting a pathological condition in a patient, for diagnostic purposes, for screening for antineoplastic agents and carcinogens, for diagnostic staging of neoplasia, for producing recombinant flap endonuclease for use as research or diagnostic reagents, for producing antibodies reactive with the novel polypeptides, for producing transgenic nonhuman animals expressing the novel polypeptides encoded by a transgene. The invention also provides novel molecular cloning techniques and reagents involving cleavage of a flap or nick with a flap endonuclease.

XX Sequence 382 AA;

Query Match 48.4%; Score 938; DB 20; Length 382;

Best Local Similarity 49.5%; Pred. No. 4.6e-78;

Matches 190; Conservative 69; Mismatches 109; Indels 16; Gaps 6;

QY 1 MGIGLTKLLADNAPKAMKEQKFESEYFGKRTAVDASMSIYQFLIVVGRGTMETLTNEAGE 60

Db 1 MGIGLNAIISEHVPSAIRKSDIKSFGRKVAIDASMSIYQFLIARQDDGQLTNEAGE 60

QY 61 VTSLQGMFNRTIRLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

Db 61 TTSLHMGFYRTLRMDNGIKPVYVFDGKPPDLKSHLTKRSSRVETEKLAETTELE 120

QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPPVDPAPSEAEACALCINDKVFVASED 180

Db 121 K---MKQERRLVKVSKHENEAQLLGLMGIPYIIAPTAEAAQCAELAKGKVVAAASD 177

QY 181 MDSLTFGAPRFLRHLMOPSSKKIPWMEFDVAKVLEELTMDQFDLCLCGCDYCDISK 240

Db 178 MDCLCYRTPFLRLHMTASEAKLPQIEFHLKVLQELGNGEQFVLDCLIGSDYCESIR 237

QY 241 GIGGOTALKLIROHGSIESILENL-----NKDRYQIPEDWPYQEARLFKEPNVTLDIPE 295

Db 238 GVGPPVATLKLIKTHGSTIEKIVEFTESGESNNTKWKIPEDWPYKQARMLFLDPEV-IDGNE 296

QY 296 --LKTADPEGLISLVKDNFGNEDRVTKALEIKSAKNSOGRLSEFFKPTATT--- 350  
 DB 297 INLKWSPPEKELEIYLCDDKKFSEERWKGISRLKGLKSGIQGRLDGFFQVVPKTRQ 356  
 QY 351 -SAPLKRKETSDKTSAANKTK 373  
 DB 357 LAAAKRAQENKLNK-NKNKTK 379

RESULT 10  
 ABG19545  
 ID ABG19545 standard; Protein; 373 AA.  
 XX  
 AC ABG19545;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #19536.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS83732.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 49904; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 373 AA;

Query Match 36.1%; Score 699.5; DB 22; Length 373;  
 Best Local Similarity 41.1%; Pred. No. 4.9e-56;

Matches 155; Conservative 68; Mismatches 105; Indels 49; Gaps 7;

QY 1 MGIGITLLADNAPKAMKEQKFESYFGKIAVDASMSYQFLIVVGRGTMETLTNEAGE 60  
 DB 36 MGIOGLAKLIADVAPSAIRENDIKSTYFGRKVAIDASMSYHFLIIV-ROGGDVLQNEEGE 94  
 QY 61 VTSHLQGMENRTIRLEAGIKFPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120  
 DB 95 TTSHLMGMFYRTIRMKENGKPIY-----KSGELAKCSELRAEAEKQLQQAAGA 145  
 QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFAYASED 180  
 DB 146 EQVVEKFTKWLKVTQKHNDRCKHLLSLMGIPYLDAPIEAEASCAALVKGKYPVATED 205  
 QY 181 MDLSLTFGAPRFLHLMDPSSKKIPVMEFDVAKVLELELTMDQFIDLCLICGGDYCDISK 240  
 DB 206 MDCFTFGSPVLMQHLTASS-----CYCKSIW 231  
 QY 241 GIGGOTALKLIHQHGSIESILENLNDRYQIPEDWPYQARRLFKEPNVT-LDIPELKWT 299  
 DB 232 SIGPKRAVDLIQKHKSIEIVQRLDPNKYPLPENRLHKEAYQLFLEPEVLGPESVELKWS 291  
 QY 300 APDEGLISFLYKDNCFNEDRVTKATEIKSAKNKSQGRLESFFKPTATTTSAPLKRKE- 358  
 DB 292 EPNEELVKFVCGEQKFSEERHSHGVKRLTEKG-GGEPRGPDWTQP--GRASPTKRPDH 348  
 QY 359 TSDKTSKAAANKTKRAG 375  
 DB 349 QEDQHTRSKSGSHXG 365

RESULT 11  
 AAW24216  
 ID AAW24216 standard; Protein; 340 AA.  
 XX  
 AC AAW24216;  
 XX  
 DT 14-APR-1998 (first entry)  
 DE Pyrococcus furiosus FEN-1 endonuclease.  
 KW Nucleic acid cleavage; 5' nuclease; DNA cleavage; RNA cleavage;  
 KW invader directed cleavage; FEN-1; endonuclease.  
 XX  
 OS Pyrococcus furiosus.  
 XX  
 PN WO9727214-A1.  
 XX  
 PD 31-JUL-1997.  
 PF 22-JAN-1997; 97WO-US01072.  
 XX  
 PR 02-DEC-1996; 96US-0759038.  
 PR 24-JAN-1996; 96US-059491.  
 PR 12-JUL-1996; 96US-0882853.  
 PR 29-NOV-1996; 96US-0756386.  
 PR 02-DEC-1996; 96US-0758314.  
 XX  
 PA (THIR-) THIRD WAVE TECHNOLOGIES INC.  
 XX  
 PI Brow MAD, Dahlberg JE, Hall JG, Kaiser MW, Lyamichev VI;  
 PI Olive DM, Prudent JR;  
 XX  
 DR WPI; 1997-393613/36.  
 DR N-PSDB; AAT76685.  
 XX  
 PT Thermostable structure-specific nuclease(s) - used for detection and  
 PT characterisation of nucleic acid sequences and variations in nucleic  
 PT acid sequences  
 XX  
 PS Example 28; Page 283-285; 457pp; English.  
 XX  
 CC This sequence comprises Pyrococcus furiosus (Pfu) FEN-1

CC endonuclease. Large-scale production of the enzyme was performed  
 CC using *E. coli* host cells transformed with a vector carrying the Pfu  
 CC FEN-1 coding sequence (see AAT76685). Pfu FEN-1 is a thermostable  
 CC enzyme. It can be used in novel methods for the detection and  
 CC characterisation of nucleic acid sequences and variations in  
 CC nucleic acid sequences.  
 XX Sequence 340 AA;  
 SQ

Query Match 33.4%; Score 648.5; DB 18; Length 340;  
 Best Local Similarity 41.3%; Pred. No. 2.2e-51;  
 Matches 142; Conservative 62; Mismatches 107; Indels 33; Gaps 6;

QY 19 KQKESYEGRIKIVDASMSIYQFLIVGRTGMEITNEAGEVTS HQGMFNRITLLEA 78  
 DB 12 KEIELENYGKKIAIDALNAIYQFLSTIRQKDGTPLMDSKGRITSHLSGLFYRTINLMEA 71

QY 79 GIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGDGDAIEKLSKRTVKVTRQH 138  
 DB 72 GIKPVYVFDGEPPEPKKELEKRRAREAEKWEAEKEIEEARKYAQATRVNEML 131

QY 139 NEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASEDMDSITFGAPRFLRLMDP 198  
 DB 132 IEDAKKLELMGIPVQAPSEGAQAAYMAAKGSVYASQDYDLSLLFGAPRLVRNLIT 191

QY 199 SSKKIP-----VMEFDVAKVLELELTMDQFIDLCILCGDY-CDSIKGIGG 244  
 DB 192 GKRKLPGKNVYVEIKPELIILE-----EVLKELKLTREKLIELAILVGYDYNPGGKIGL 247

QY 245 QTALKLRHQHSIESILENLKND---RYQIPEDWPYQEARLKEPNVTLDIPELKWTP 301  
 DB 248 KKALEIVRH-----SKDPLAKFKQSDVDLYAIKEFFLNPPVT-DNYNLVWRDP 295

QY 302 DEEGLISFLVNDGNFEDRVTKAEIKSAKNKSSQGRLESFFK 345  
 DB 296 DEEGLIKFLCDEHDFSEERVKNGLERLKAIAKSKQSTLESWFK 339

RESULT 12  
 AAW79970  
 ID AAW79970 standard; Protein; 340 AA.  
 AC AAW79970;  
 XX  
 XX  
 DT 02-FEB-1999 (first entry)  
 XX  
 XX  
 DE Pyrococcus furiosus FEN-1 endonuclease.  
 KW Nucleic acid detection; multiple sequential invasive cleavage;  
 KW FEN-1; endonuclease; nuclease.  
 XX  
 OS Pyrococcus furiosus.  
 XX  
 PN WO9842873-A1.  
 XX  
 PD 01-OCT-1998.  
 XX  
 PF 24-MAR-1998; 98WO-USO5809.  
 XX  
 PR 24-MAR-1997; 97US-0823516.  
 XX  
 XX (THIR-) THIRD WAVE TECHNOLOGIES INC.  
 XX  
 XX BroW MAD, Hall JG, Kwiatkowski RW, Lyamichev VI;  
 PI Mast AL, Vavra SH;  
 PI  
 XX WPI; 1998-557036/47.  
 DR N-PSDB; AAV65840.  
 XX  
 XX Detecting target nucleic acid by sequence-specific cleavage of  
 PT complex with two specific oligonucleotides - used to detect  
 PT cytomegalovirus DNA  
 XX

PS  
 XX  
 CC This is the amino acid sequence of FEN-1 endonuclease of  
 CC Pyrococcus furiosus (Pfu). FEN-1 DNA (see AAV65840) has been  
 CC ligated into vector for expression in *E. coli* cells. The invention  
 CC relates to means for the detection and characterisation of nucleic  
 CC acid sequences, and variations in nucleic acid sequences. It also  
 CC relates to methods for forming a nucleic acid cleavage structure on  
 CC a target sequence and cleaving this structure in a site-specific  
 CC manner, preferably using a thermostable structure-specific nuclease  
 CC such as FEN-1. Cleavage of the cleavage structure by the  
 CC nuclease indicates the presence of specific nucleic acid sequences  
 CC or specific variants. The invention further relates to methods for  
 CC the separation of nucleic acid molecules based on charge, methods  
 CC for the detection of non-target cleavage products via the formation  
 CC of a complete and activated protein binding region, and methods for  
 CC the detection of nucleic acid from various viruses (e.g. human  
 CC cytomegalovirus) in a sample. The method amplifies the detection  
 CC molecule rather than the target itself, is less subject to  
 CC contamination than exponential amplification processes, and allows  
 CC many targets to be analysed in a single reaction.  
 XX Sequence 340 AA;  
 SQ

Query Match 33.4%; Score 648.5; DB 19; Length 340;  
 Best Local Similarity 41.3%; Pred. No. 2.2e-51;  
 Matches 142; Conservative 62; Mismatches 107; Indels 33; Gaps 6;

QY 19 KQKESYEGRIKIVDASMSIYQFLIVGRTGMEITNEAGEVTS HQGMFNRITLLEA 78  
 DB 12 KEIELENYGKKIAIDALNAIYQFLSTIRQKDGTPLMDSKGRITSHLSGLFYRTINLMEA 71

QY 79 GIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGDGDAIEKLSKRTVKVTRQH 138  
 DB 72 GIKPVYVFDGEPPEPKKELEKRRAREAEKWEAEKEIEEARKYAQATRVNEML 131

QY 139 NEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASEDMDSITFGAPRFLRLMDP 198  
 DB 132 IEDAKKLELMGIPVQAPSEGAQAAYMAAKGSVYASQDYDLSLLFGAPRLVRNLIT 191

QY 199 SSKKIP-----VMEFDVAKVLELELTMDQFIDLCILCGDY-CDSIKGIGG 244  
 DB 192 GKRKLPGKNVYVEIKPELIILE-----EVLKELKLTREKLIELAILVGYDYNPGGKIGL 247

QY 245 QTALKLRHQHSIESILENLKND---RYQIPEDWPYQEARLKEPNVTLDIPELKWTP 301  
 DB 248 KKALEIVRH-----SKDPLAKFKQSDVDLYAIKEFFLNPPVT-DNYNLVWRDP 295

QY 302 DEEGLISFLVNDGNFEDRVTKAEIKSAKNKSSQGRLESFFK 345  
 DB 296 DEEGLIKFLCDEHDFSEERVKNGLERLKAIAKSKQSTLESWFK 339

RESULT 13  
 AAW59940  
 ID AAW59940 standard; Protein; 340 AA.  
 AC AAW59940;  
 XX  
 XX  
 DT 21-DEC-1998 (first entry)  
 XX  
 XX Amino acid sequence of the Pfu FEN-1 endonuclease.  
 XX Mja FEN-1 endonuclease; Taq gene; structure-specific nuclease;  
 KW mutant; DNA polymerase; bacteria; fungi; protozoa; RNA virus;  
 KW hepatitis C virus; HCV; thermostable.  
 XX  
 OS Pyrococcus furiosus.  
 XX  
 PN WO9823774-A1.  
 XX  
 PD 04-JUN-1998.  
 XX

26-NOV-1997; 97WO-US21783.

02-DEC-1996; 96US-0758314.  
29-NOV-1996; 96US-0757653.

(THIR-) THIRD WAVE TECHNOLOGIES INC.

Kaiser MW, Lyamichev VI, Lyamicheva N;  
WPI; 1998-322748/28.  
N-PSDB; AAV53951.

Thermostable structure-specific nuclease(s) derived from mutant DNA polymerase(s) - useful for detecting mutant allele(s) or strains of microorganisms

Example 29; Pages 280-281; 472pp; English.

This is the amino acid sequence encoding the Pfu FEN-1 endonuclease, used in the method of the invention. In this process thermostable structure-specific nucleases are derived from mutant DNA polymerases, which can be used for detecting mutant alleles or strains of microorganisms. The structure-specific nucleases can be used in mixtures, compositions and kits to treat nucleic acid, e.g. for detection of wild type and mutant alleles of genes, for detection and/or identification of strains of microorganisms such as bacteria, fungi, protozoa, especially for detection of RNA viruses such as the hepatitis C virus (HCV).

Sequence 340 AA;

Query Match 33.4%; Score 648.5; DB 19; Length 340;  
Best Local Similarity 41.3%; Pred. No. 2.2e-51;  
Matches 142; Conservative 62; Mismatches 107; Indels 33; Gaps

Qy 19 KEQFESYFGKTAVDASMSIYQFLIVGRTGTMETLTNEAGEVTSHLQGMENRTIRLEA 78  
|| : : : || : || : || : || : : : : : : || : || : || : || : || : || : ||  
Db 12 KEIELENYGKKIADALNAIYQFLSTIRKOGTPTLMDSKGRITSHLSGLFRTINLMEA 71  
|| : : : || : || : || : || : : : : : : || : || : || : || : || : || : ||  
Qy 79 GIKPVYVFDGKPPDMKQOELAKRYSKRDDATKDLTEAVEVGDKDAIEKLRSKRTVKYTRQ 138  
|| : : : || : || : || : || : : : : : : || : || : : : : : : : : : :  
Db 72 GIKPVYVFDGPEPFKKKELEKREAREAEAEKWRALEKGEIEEARKYAQRATRVNEML 131  
|| : : : || : || : || : || : : : : : : || : || : : : : : : : : : :  
Qy 139 NEDCKRLRLMGVPEVPAPSEAEACALCINDKVFASSEDMSLTFGAPRFLRHLMDP 198  
|| : : : || : || : || : || : : : : : : || : || : || : || : || : || : ||  
Db 132 IEDAKLLELMGIPVQAPSEGEQAAYMAKGSVVASASQDYDLSLLFGAPRLVRNLTT 191  
|| : : : || : || : || : || : : : : : : || : || : || : || : || : || : ||  
Qy 199 SSKIP-----VMEFDVAKVLEELTMDQFDILCILCGCY-CDSIKGIGG 244  
:  
Db 192 GKRLPGKNVVEIKPELIIILE---EVLEKELTKREKLIELAILVGTDPNGGIKIGL 247  
:  
Qy 245 QATALKLRQHSIESILENLKND---RYQIPEDWPYQEARRLFKEPNVTIDTIDPELKWTP 301  
:  
Db 248 KKALEIVRH-----SKDPLAKFQKQSDVDLYAKPEFLNPVT-DYNLVRDP 295  
|| : : : || : || : || : || : : : : : : || : || : || : || : || : || : ||  
Qy 302 DEGLISFLVKDNGFNEDRVTKAIEKISAKNKSQGRLESFFK 345  
|| : : : || : || : || : || : : : : : : || : || : || : || : || : || : ||  
Db 296 DEGLIAKFLCDEHDFSEERVKNGLERLKAIKSGKQSTLESWFK 339  
|| : : : || : || : || : || : : : : : : || : || : || : || : || : || : ||

RESULT 14  
AAW59953  
ID AAW59953 standard; Protein; 326 AA.  
AC AAW59953;  
XX AAW59953;  
XX  
22-DEC-1998 (first entry)  
XX  
DE Amino acid sequence of the structure specific nuclease 5.  
XX  
KW Taq mutant gene; thermostable; nuclease; mutant; DNA polymerase;  
KW bacteria; protozoa; fungi; RNA virus; hepatitis C virus; HCV.  
XX  
XX

DE Amino acid sequence of the structure specific nuclease 3.  
 KW Taq mutant gene; thermostable; nuclease; mutant; DNA polymerase;  
 KW bacteria; protozoa; fungi; RNA virus; hepatitis C virus; HCV.  
 XX  
 XX Synthetic.  
 OS  
 XX  
 XX WO9823774-A1.  
 PN  
 XX  
 XX 04-JUN-1998.  
 PD  
 XX  
 XX 26-NOV-1997; 97WO-US21783.  
 XX  
 XX  
 XX 02-DEC-1996; 96US-0758314.  
 PR  
 XX 29-NOV-1996; 96US-0757653.  
 PR  
 XX  
 XX (THIR-) THIRD WAVE TECHNOLOGIES INC.  
 PA  
 XX  
 XX Kaiser MW, Lyamichev VI, Lyamicheva N;  
 PI  
 XX  
 XX WPI; 1998-322748/28.  
 DR  
 XX  
 XX Thermostable structure-specific nuclease(s) derived from mutant DNA  
 PT polymerase(s) - useful for detecting mutant allele(s) or strains of  
 PT microorganisms  
 PT  
 XX  
 XX Claim 1; Pages 329-330; 472pp; English.  
 PS  
 XX  
 XX This is the amino acid sequence of a structure specific nuclease  
 CC used in the method of the invention. In this process thermostable  
 CC structure-specific nucleases are derived from mutant DNA polymerases,  
 CC which can be used for detecting mutant alleles or strains of  
 CC microorganisms. The structure-specific nucleases can be used in  
 CC mixtures, compositions and kits to treat nucleic acid, e.g. for  
 CC detection of wild type and mutant alleles of genes, for detection  
 CC and/or identification of strains of microorganisms such as bacteria,  
 CC fungi, protozoa, especially for detection of RNA viruses such as the  
 CC hepatitis C virus (HCV).  
 CC  
 XX  
 XX Sequence 332 AA;  
 SQ

Query Match 33.2%; Score 644; DB 19; Length 332;  
 Best Local Similarity 43.2%; Pred. No. 5.6e-51;  
 Matches 142; Conservative 68; Mismatches 103; Indels 16; Gaps 7;

QY 19 KEOKFESYFGKTIADVDSMSIYQFLIVVGRGTGMEVTLTNEAGEVTSHLQGMFNTIRLEA 78  
 Db 12 KELENLGKKTADALNAYIQLSTIRQKDTPLMDSKGRITSHLSGLFYRTINLMEA 71

QY 79 GIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGDKDATEKLSKRTVKVTRQH 138  
 Db 72 GIKPVYVFDGEPPEFKKELEKREAREAEKREALEKEGEIEEAAYAKRVSYLTPKM 131

QY 139 NEDCKRLRLMGVVPVPEASEAEACALCINDKVFVASEDMDSLTFFGAPRFLRLMDP 198  
 Db 132 VENCKYLLSLMGIPYVEAPSEGEAQASNAKKGDMVAVSQDYDALLYGAPRVVRNL--T 189

QY 199 SSKKIPVMEFDVAKVLEELTMDQFDLCLCGCDY-CDSIKGIGGQATKLIRHGS 257  
 Db 190 TTKEMPEL-TELNEVLDELRLISLDDIDIAIFMGTFNPGGVKYGIGIKRAYELVRS-GVA 247

QY 258 ESILENLNDRYQIPEDWPYQEARLFKPNVTLDIPELKWTAPDEGLISFLVKDNGFN 317  
 Db 248 KDVLLK-----EVEYYDEIKRIFKPKVT-DNYSLSKLDPKREGIIFLVNDENFN 297

QY 318 EDRVTKAIEKIKS-AKNKSSQGRLESFFK 345  
 Db 298 YDRVKKHVDKLYNLANKTKQKTLDAWEK 326

Search completed: May 8, 2003, 21:40:45  
 Job time : 64 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:39:38 ; Search time 34 Seconds  
(without alignments)  
327.979 Million cell updates/sec

Title: US-09-805-311-2

Perfect score: 1939

Sequence: 1 MGKGLTKLLADNAPKAMKE.....SDKTSKAAANKTKAGGKKK 379

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfilesi.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1939	100.0	379	4	US-09-426-557-2	Sequence 2, Appli
2	1939	100.0	379	4	US-09-426-557-6	Sequence 6, Appli
3	1933	99.7	379	4	US-09-426-557-4	Sequence 4, Appli
4	1933	99.7	379	4	US-09-426-557-8	Sequence 8, Appli
5	1032	53.2	380	2	US-08-823-516-137	Sequence 137, App
6	1032	53.2	380	4	US-09-426-557-10	Sequence 10, Appli
7	1031	53.2	380	2	US-08-455-968E-1	Sequence 1, Appli
8	1014	52.3	378	2	US-08-823-516-138	Sequence 138, App
9	1007.5	52.0	377	2	US-08-455-968E-3	Sequence 3, Appli
10	996	51.4	378	2	US-08-455-968E-10	Sequence 10, Appli
11	938	48.4	382	2	US-08-455-968E-5	Sequence 5, Appli
12	938	48.4	382	2	US-08-823-516-139	Sequence 139, App
13	648.5	33.4	340	2	US-08-757-653-176	Sequence 176, App
14	648.5	33.4	340	2	US-08-823-516-79	Sequence 79, Appli
15	648.5	33.4	340	2	US-08-823-516-136	Sequence 136, App
16	648.5	33.4	340	3	US-08-759-038-115	Sequence 115, App
17	648.5	33.4	340	3	US-08-314-115	Sequence 115, App
18	632.5	32.6	343	4	US-09-146-319-2	Sequence 2, Appli
19	632.5	32.6	343	4	US-09-175-973-2	Sequence 2, Appli
20	560.5	28.9	326	2	US-08-757-653-172	Sequence 172, App
21	560.5	28.9	326	2	US-08-823-516-75	Sequence 75, Appli
22	560.5	28.9	326	2	US-08-823-516-135	Sequence 135, App
23	560.5	28.9	326	3	US-08-759-038-111	Sequence 111, App
24	560.5	28.9	326	3	US-08-314-111	Sequence 111, App
25	364	18.8	387	2	US-08-823-516-140	Sequence 140, App
26	352.5	18.2	488	2	US-08-823-516-141	Sequence 141, App
27	340.5	17.6	386	2	US-08-455-968E-7	Sequence 7, Appli

28	340	17.5	543	2	US-08-823-516-143	Sequence 143, App
29	334.5	17.3	550	2	US-08-823-516-142	Sequence 142, App
30	313.5	16.2	527	2	US-08-823-516-144	Sequence 144, App
31	205	10.6	872	1	US-08-766-014-2	Sequence 2, Appli
32	184.5	9.5	291	4	US-09-105-697-6	Sequence 6, Appli
33	173.5	8.9	834	5	PCT-US95-14418-2	Sequence 2, Appli
34	173.5	8.9	834	5	PCT-US95-15327-2	Sequence 2, Appli
35	173	8.9	548	2	US-08-484-956-86	Sequence 86, Appli
36	173	8.9	548	2	US-08-757-653-86	Sequence 86, Appli
37	173	8.9	548	2	US-08-520-946-86	Sequence 86, Appli
38	173	8.9	695	2	US-08-484-956-87	Sequence 87, Appli
39	173	8.9	695	2	US-08-757-653-87	Sequence 87, Appli
40	173	8.9	695	4	US-08-520-946-87	Sequence 87, Appli
41	173	8.9	810	4	US-09-587-856-2	Sequence 2, Appli
42	173	8.9	810	4	US-09-777-537-2	Sequence 2, Appli
43	173	8.9	810	4	US-09-777-538-2	Sequence 2, Appli
44	173	8.9	832	1	US-07-977-434-2	Sequence 2, Appli
45	173	8.9	832	1	US-08-156-020-2	Sequence 2, Appli

## ALIGNMENTS

### RESULT 1

US-09-426-557-2

; Sequence 2, Application US/09426557

; Patent No. 6232527

; GENERAL INFORMATION:

; APPLICANT: Mahajan, Pramod B.

; TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses

; FILE REFERENCE: Thereof

; CURRENT APPLICATION NUMBER: US/09/426,557

; CURRENT FILING DATE: 1999-10-22

; EARLIER APPLICATION NUMBER: 60/112,332

; EARLIER FILING DATE: 1998-12-15

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 379

; TYPE: PRT

; ORGANISM: Zea mays

US-09-426-557-2

Query Match 100.0%; Score 1939; DB 4; Length 379;  
Best Local Similarity 100.0%; Pred No. 1.2e-183;  
Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGKGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIVGRTGTMETLTNEAGE	60
Db	1	MGKGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIVGRTGTMETLTNEAGE	60
QY	61	VTSHLOGMNRTRIRLEAGIKPVYVFDGKPPDMKKQBLAKRYSKRDATAKDLTAVEVGD	120
Db	61	VTSHLOGMNRTRIRLEAGIKPVYVFDGKPPDMKKQBLAKRYSKRDATAKDLTAVEVGD	120
QY	121	KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPPVVEAPSEAECAALCINDKVFVASED	180
Db	121	KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPPVVEAPSEAECAALCINDKVFVASED	180
QY	181	MDSLTFGAPRFLRHLMDPSSKIPVMFEFVAKVLEELTMDQFIDLCIICGCTCDSIK	240
Db	181	MDSLTFGAPRFLRHLMDPSSKIPVMFEFVAKVLEELTMDQFIDLCIICGCTCDSIK	240
QY	241	GIGGQTALRIHQHSIESILENLNDRYQIPEDWPQEARRLFKEPNVTLDIPELKWTA	300
Db	241	GIGGQTALRIHQHSIESILENLNDRYQIPEDWPQEARRLFKEPNVTLDIPELKWTA	300
QY	301	PDEEGLISFLVDNGFNEDRVTKAIEKISAKNKSOGRLSEFPKPTATTAPLKRKETS	360
Db	301	PDEEGLISFLVDNGFNEDRVTKAIEKISAKNKSOGRLSEFPKPTATTAPLKRKETS	360
QY	361	DKTSKAAANKTKAGGKKK	379





Db 121 KDAEKLKSKRTVKTROHNECDKRLRLMGVPVVEAPSEAECAALCINDKVFVASED 180  
QY 181 MDSLTGAPRFLRHLMDPSSKKIPVMEFDVAKVLELELMDQFIDICILCGGDCYDSIK 240  
Db 181 KDSLTGAPRFLRHLMDPSSKKIPVMEFDVAKVLELELMDQFIDICILCGGDCYDSIK 240  
QY 241 GIGQTALKLIRHGSTESILENLDKRYOIPEDWPYQEARRLFKEPNVTLDIPELKWTA 300  
Db 241 GIGQTALKLIRHGSTESILENLDKRYOIPEDWPYQEARRLFKEPNVTLDIPELKWTA 300  
QY 301 PDEEGLISFLVKGNGFNEEDRVTKAIEKISAKNKSOGRLSEFFKPTATTSAPLKRKETS 360  
Db 301 PDEEGLISFLVKGNGFNEEDRVTKAIEKISAKNKSOGRLSEFFKPTATTSAPLKRKETS 360  
QY 361 DKTSKAAANKTKRAGGKK 379  
Db 361 DKTSKAAANKTKRAGGKK 379

## RESULT 5

US-0823-516-137  
; Sequence 137, Application US/0823516  
; Patent No. 5994069  
; GENERAL INFORMATION:  
; APPLICANT: Hall, Jeff G.  
; APPLICANT: Lyamichev, Victor I.  
; APPLICANT: Mast, Andrea L.  
; APPLICANT: Brow, Mary Ann D.  
; TITLE OF INVENTION: Detection of Nucleic Acids By Multiple  
; TITLE OF INVENTION: Sequential Invasive Cleavages  
; NUMBER OF SEQUENCES: 163  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlin & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/823,516  
; FILING DATE: 24-MAR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/01072  
; FILING DATE: 21-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/759,038  
; FILING DATE: 02-DEC-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/758,314  
; FILING DATE: 02-DEC-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/756,386  
; FILING DATE: 29-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/682,853  
; FILING DATE: 12-JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/599,491  
; FILING DATE: 24-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: FORS-02736  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 137:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
US-0823-516-137

Query Match 53.2%; Score 1032; DB 2; Length 380;  
Best Local Similarity 53.8%; Pred. No. 7.1e-94;  
Matches 206; Conservative 69; Mismatches 98; Indels 10; Gaps 6;  
QY 1 MGIGLTKLLADNAPKAMKEQKESYFGKRIADVADSMYSIQFLIVVGRGTMTTNEAGE 60  
Db 1 MGIGLTKLLADNAPKAMKEQKESYFGKRIADVADSMYSIQFLIVVGRGTMTTNEAGE 60  
QY 61 VTSHLQGMFNRTIRLLEAGIKPVYFDGKPPDMKQELAKRYSKRDDATKDLTAEVGD 120  
Db 60 TTSMLMGMYRTIRMMENGIRPVYFDGKPPDMKQELAKRYSKRDDATKDLTAEVGD 119  
QY 121 KDAEKLKSKRTVKTROHNECDKRLRLMGVPVVEAPSEAECAALCINDKVFVASED 180  
Db 120 EQVEKFTKRLVKVTKOHNECDKRLRLMGVPVVEAPSEAECAALCINDKVFVASED 179  
QY 181 MDSLTGAPRFLRHLMDPSSKKIPVMEFDVAKVLELELMDQFIDICILCGGDCYDSIK 240  
Db 180 MDCITFGSPVLMRHLTASEAKKLPQEFHLSRIQLGELGNOEQFVDCILGSDYCESIR 239  
QY 241 GIGQTALKLIRHGSTESILENLDKRYOIPEDWPYQEARRLFKEPNVTLDIPELKW 298  
Db 240 GIGPKRAVDLIQKHKSIEEIVRRDPNKPYPENWLHKEAHLFEPEV-LDPESVELKW 298  
QY 299 TAPDEEGLISFLVKGNGFNEEDRVTKAIEKISAKNKSOGRLSEFFKPTATTSAPLKRKE 358  
Db 299 SEPNEEELIKFMCGEKQFSEERISRGVARLSKSGTQGRDLDDFFKVTGSLSS-AKRKE 357  
QY 359 TSDK--TSKAAANKTKRAGGKK 379  
Db 358 PEPKSGSTKKA--KTGAAGKFK 377

## RESULT 6

US-09-426-557-10  
; Sequence 10, Application US/09426557  
; Patent No. 6232527  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses  
; TITLE OF INVENTION: Thereof  
; FILE REFERENCE: 0961  
; CURRENT APPLICATION NUMBER: US/09/426,557  
; CURRENT FILING DATE: 1999-10-22  
; EARLIER APPLICATION NUMBER: 60/112,332  
; EARLIER FILING DATE: 1998-12-15  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-426-557-10

Query Match 53.2%; Score 1032; DB 4; Length 380;  
Best Local Similarity 53.8%; Pred. No. 7.1e-94;  
Matches 206; Conservative 69; Mismatches 98; Indels 10; Gaps 6;  
QY 1 MGIGLTKLLADNAPKAMKEQKESYFGKRIADVADSMYSIQFLIVVGRGTMTTNEAGE 60  
Db 1 MGIGLTKLLADNAPKAMKEQKESYFGKRIADVADSMYSIQFLIVVGRGTMTTNEAGE 59  
QY 61 VTSHLQGMFNRTIRLLEAGIKPVYFDGKPPDMKQELAKRYSKRDDATKDLTAEVGD 120  
Db 60 TTSMLMGMYRTIRMMENGIRPVYFDGKPPDMKQELAKRYSKRDDATKDLTAEVGD 119

Db 60 TTSHLMGFYRTIRMENCIKPVYVDFDGKPPOLKSGELAKRSERRAEAEKQLQQAAGA 119  
 QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPPVVEAPSEAEAEALCINDKVFVASED 180  
 Db 120 EQEVEKFTKRLVKVTKQHNDCKHLLSLMGIPYLDAPSEAEAEALCINDKVFVASED 179  
 QY 181 MDSLTFGAPRFLRHLMDPSSKIPVMEFDVAKVLELELTMDQFDLCLCGCCDYCDSTK 240  
 Db 180 MDCLTFGSPVLMRHLTASEAKLPQIEFHLRSLQELGNOEQFDVLDCLILGSDYCESIR 239  
 QY 241 GIGGOTALKLRHQHSIESILENKNDRYQIPEDWYOEARLFKEPNVTLD--IPELKW 298  
 Db 240 GIGKPRVDLQKHKSIEIIVRLDPNKPYPENWLHKEAHLFLEPEV-LDPESVELKW 298  
 QY 299 TAPDEGLISLVDKNGNEDRVTKAIEKISAKNKSQGRLESFFKPTATTSAPLKRKE 358  
 Db 299 SEPNEELIKFMCGEKQSEERIRSGVRLSKSRQSGTQGRDLDFFKVTGSLSS-AKRKE 357  
 QY 359 TSDK--TSKAAANKTKAGGKK 379  
 Db 358 PEPKSTKKA---KTGAAGKFK 377

## RESULT 7

US-08-455-968E-1

; Sequence 1, Application US/08455968E

; Patent No. 5874283

; GENERAL INFORMATION:

; APPLICANT: Harrington, John L.

; APPLICANT: Hsieh, Chih-Lin

; APPLICANT: Lieber, Michael

; TITLE OF INVENTION: Mammalian Flap-Specific Endonuclease

; NUMBER OF SEQUENCES: 63

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/455,968E

; FILING DATE: 30-MAY-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Apple, Randolph T.

; REGISTRATION NUMBER: 36,429

; REFERENCE/DOCKET NUMBER: 18985-000100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 380 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-455-968E-1

Query Match

Best Local Similarity 53.2%; Score 1031; DB 2; Length 380;

Matches 206; Conservative 69; Mismatches 98; Indels 10; Gaps 6;

QY 1 MGIGKGLKLLADNAPKAMKQKFSYGRKIADVDSMSIQFLIVGRTGNETLTNAGE 60

Db 1 MGIGKGLKLLADNAPKAMKQKFSYGRKIADVDSMSIQFLIVGRTGNETLTNAGE 59

QY 61 VTSHLQGMFNRTIRLERAGIKPVYVDFDGKPPDMKKOEALAKRSKRDDATKDLTEAVEVGD 120  
 Db 60 TTSHLMGFYRTIRMENCIKPVYVDFDGKPPOLKSGELAKRSERRAEAEKQLQQAAGA 119  
 QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPPVVEAPSEAEAEALCINDKVFVASED 180  
 Db 120 EQEVEKFTKRLVKVTKQHNDCKHLLSLMGIPYLDAPSEAEAEALCINDKVFVASED 179  
 QY 181 MDSLTFGAPRFLRHLMDPSSKIPVMEFDVAKVLELELTMDQFDLCLCGCCDYCDSTK 240  
 Db 180 MDCLTFGSPVLMRHLTASEAKLPQIEFHLRSLQELGNOEQFDVLDCLILGSDYCESIR 239  
 QY 241 GIGGOTALKLRHQHSIESILENKNDRYQIPEDWYOEARLFKEPNVTLD--IPELKW 298  
 Db 240 GIGKPRVDLQKHKSIEIIVRLDPNKPYPENWLHKEAHLFLEPEV-LDPESVELKW 298  
 QY 299 TAPDEGLISLVDKNGNEDRVTKAIEKISAKNKSQGRLESFFKPTATTSAPLKRKE 358  
 Db 299 SEPNEELIKFMCGEKQSEERIRSGVRLSKSRQSGTQGRDLDFFKVTGSLSS-AKRKE 357  
 QY 359 TSDK--TSKAAANKTKAGGKK 379  
 Db 358 PEPKSTKKA---KTGAAGKFK 377

## RESULT 8

US-08-823-516-138

; Sequence 138, Application US/08823516

; Patent No. 5994069

; GENERAL INFORMATION:

; APPLICANT: Hall, Jeff G.

; APPLICANT: Lyamichev, Victor I.

; APPLICANT: Mast, Andrea L.

; APPLICANT: Brow, Mary Ann D.

; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple

; TITLE OF INVENTION: Sequential Invasive Cleavages

; NUMBER OF SEQUENCES: 163

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Medlen &amp; Carroll, LLP

; STREET: 220 Montgomery Street, Suite 2200

; CITY: San Francisco

; STATE: California

; COUNTRY: United States Of America

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/823,516

; FILING DATE: 24-MAR-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US97/01072

; FILING DATE: 21-JAN-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/759,038

; FILING DATE: 02-DEC-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/758,314

; FILING DATE: 02-DEC-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/756,386

; FILING DATE: 29-NOV-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/682,853

; FILING DATE: 12-JUL-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/599,491

; FILING DATE: 24-JAN-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Ingolia, Diane E.

CLASSIFICATION: 435



```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: FILING DATE: 24-MAR-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US97/01072
: FILING DATE: 21-JAN-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/759,038
: FILING DATE: 02-DEC-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/758,314
: FILING DATE: 02-DEC-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/756,386
: FILING DATE: 29-NOV-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/682,853
: FILING DATE: 12-JUL-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/599,491
: FILING DATE: 24-JAN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Ingolia, Diane E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: FORS-02736
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 139:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 382 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: US-08-823-516-139

```

```

Query Match 48.4%; Score 938; DB 2; Length 382;
Best Local Similarity 49.5%; Pred. No. 1.4e-84;
Matches 190; Conservative 69; Mismatches 109; Indels 16; Gaps 6;

Qy 1 MGKGLTKLLADNAPKAMEQEFYGRKTIADVADSMISYQFLIVVGTGNETLTNEAGE 60
Db 1 MGKGLNAAIISFVPSAIRKSDIKSFGRKVAIDASMSLYQFLIATVQQDGGQLTNEAGE 60

Qy 61 VTSHLQGMENRTIRLLEAGIKPVYFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120
Db 61 TTSHLGMEYRTLRMIDNGIKPCYVFDGKPPDLKSHLTKSSRRVETKKLAETTELE 120

Qy 121 KDATELKSRTKVTATROHDECKRLRLMGVPEAPSEAECAALCINDKVFVASED 180
Db 121 K--MQERLVVKSHEEAAQLGLMGIPYITAPTEAQAQCAELAKGKGVYAAASED 177

Qy 181 MDSITFGAPRLHLMDSKPIVMEFDVAKVLEELTMDQFTDLCILCGDCYDSIK 240
Db 178 MDITCYTPPRLRLHFTFAKKEPIHEIDTLVLRGLDITIEQFVDCILMGDCYCSIR 237

Qy 241 GIGQOTALKIRHGSIESILNENL-----NKDRYQIPEDWPYQEARRLFKEPNVTLDIPE 295
Db 238 GVGPTALKIKTHGSIEKIVERFIESGESNNTKKWIPEDWPYQEARRLFKEPNVTLDIPE 296

Qy 296 --LKWTPADEGLISFLVNDGFENEDRVTKAIEKISAKNKSQGRLESFFKPTATT--- 350
Db 297 INLWSPKKEILYELCDKHFSEERVKSGISRLKGLKSGIOGRULDGFQVVPKTKQ 356

Qy 351 -SAPLKRKETSDKTSAKAAKTK 373
Db 357 LAAAKRAQENKKLNK-NKNKVTK 379

```

```

RESULT 13
US-08-757-653-176
: Sequence 176, Application US/08757653
: Patent No. 5843669
: GENERAL INFORMATION:
: APPLICANT: Kaiser, Michael W.
: APPLICANT: Lyamichev, Victor I.
: APPLICANT: Lyamichev, Natasha
: TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
: TITLE OF INVENTION: Thermostable FEN-1 Endonucleases
: NUMBER OF SEQUENCES: 190
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Medlen & Carroll, LLP
: STREET: 220 Montgomery Street, Suite 2200
: CITY: San Francisco
: STATE: California
: COUNTRY: United States Of America
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/757,653
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Ingolia, Diane E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: FORS-02565
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 176:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 340 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-757-653-176

Query Match 33.4%; Score 648.5; DB 2; Length 340;
Best Local Similarity 41.3%; Pred. No. 5.4e-56;
Matches 142; Conservative 62; Mismatches 107; Indels 33; Gaps 6;

Qy 19 KEQFESYFGKTIADVADSMISYQFLIVVGTGNETLTNEAGEVTSHLQGMFNRTIRLEA 78
Db 12 KEIELENLGKTIADALNAIYQFLSTIRQKDGTPLMDSKGRITSHLSGLFYRTINLMEA 71

Qy 79 GIKPVYFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGDKDAIEKLKRTVKVTRQH 138
Db 72 GIKPVYFDGPEPEFKKLEKREAREAEKWEALEKGEIEARKYAQRATFVNEML 131

Qy 139 NEDCKRLRLMGVPEAPSEAECAALCINDKVFVASEDMSLTGAPRFLHLMDP 198
Db 132 IEDAKLLELMGIPVQAPSEGAQAAYMAAKSGSVYASQDYDLSLFGAPRLVRLNLTIT 191

Qy 199 SSRKIP-----VMEFDVAKVLEELTMDQFTDLCILCGDY-CDSIKGIGG 244
Db 192 GKRLPGKNVVEITKPELITL-----EVLKELKTRKLIELAILVGTDPNGIKIGL 247

Qy 245 OTALKLIRHGSIESILENLNDK-----RYOIPEDWPYQEARRLFKEPNVTLDIPELKWTP 301
Db 248 KKALEIVRH-----SKDPLAKFQKQSDVDLYAIKEFFLNPPVT-DYNNLVWRDP 295

Qy 302 DEGLISFLVNDGFENEDRVTKAIEKISAKNKSQGRLESFFK 345
Db 296 DEEGILKFLCDEHDFSEERVKNGLERLKAIRKAKSGKOSTLESWFK 339

```

RESULT 14



```

; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-823-516-136

Query Match      33.4%; Score 648.5; DB 2; Length 340;
Best local Similarity 41.3%; Pred. No. 5.4e-56;
Matches 142; Conservative 62; Mismatches 107; Indels 33; Gaps 6;

Qy 19 KEQFESYGRKIAVDASMSIYQFLIVVGRGTMETLTNEAGEVTSHLQGMFNRTIRLLEA 78
Db 12 KEIENLYGKKIAIDALNAIYQFLSTIRQDGTPLMDSKGRITSHLSGLEFVRTINLMEA 71

Qy 79 GIRPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGGDKDALEKLSKRTVKVTRQH 138
Db 72 GIRPVYVFDGEPPEPKKKELEKREAREEAEEKWREALEKEGELEARKYAQRATRVNEML 131

Qy 139 NEDCKRLRLMGVVPVVEAPSEAEACALCINDKVFVASEDMDSLTTFGAPREFLEHMDP 198
Db 132 IEDAKKLELMGPIVQAPSGEQAQAYMAAKGSVYASASQDDYDLSLFGAPRLVRLNLTIT 191

Qy 199 SSKKIP-----VMEFDVAKVLELELTMDQFDLCLCGDGY-CDSIKGIGG 244
Db 192 GKKLFCKNYYVEIKPELTILE---EVLKELKLTREKLELAILVGTDPNGGKIGIGL 247

Qy 245 QTALKLIRHGSTESILENLKRD---RYQIPEDWPYQEARRLFKEPNVTLDIPELKWTPAP 301
Db 248 KKALETVRH-----SKDPLAKFQKQSDVDLYAIKEFFLNPPVT-DNYNLVWRDP 295

Qy 302 DEEGLISFLVKONGNEDRVTKATEKTKSAKNKSSOGRLESEFK 345
Db 296 DEEGILKFLCDEHDFSEERVKNGLERLKAISKQKSTLESWPK 339
```

Search completed: May 8, 2003, 21:44:30  
Job time : 35 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:40:52 ; Search time 58 Seconds  
(without alignments)  
601.340 Million cell updates/sec

Title: US-09-805-311-2

Perfect score: 1939

Sequence: 1 MGKGLTKLLADNAPKAME.....SDRTSKAAANKTKAGRKK 379

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB\_PEP.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB\_PEP.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB\_PEP.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB\_PEP.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB\_PEP.\*
- 6: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB\_PEP.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB\_PEP.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB\_PEP.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB\_PEP.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09\_PUBCOMB\_PEP.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB\_PEP.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US10\_PUBCOMB\_PEP.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB\_PEP.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB\_PEP.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1939	100.0	379	10	US-09-805-311-2
2	1939	100.0	379	10	US-09-805-311-2
3	1933	99.7	379	10	US-09-805-311-4
4	1933	99.7	379	10	US-09-805-311-8
5	1032	53.2	380	9	US-10-033-297-137
6	1032	53.2	380	9	US-09-940-244-137
7	1032	53.2	380	10	US-09-805-311-10
8	1014	52.3	378	9	US-10-033-297-138
9	1014	52.3	378	9	US-09-940-244-138
10	938	48.4	382	9	US-09-940-244-139
11	938	48.4	382	9	US-10-033-297-139
12	648.5	33.4	340	9	US-10-033-297-136
13	648.5	33.4	340	9	US-09-940-244-136
14	648.5	33.4	340	9	US-09-940-244-136
15	648.5	33.4	340	10	US-09-777-430A-27
16	637	32.6	340	9	US-09-940-244-365
17	632.5	32.9	343	9	US-09-940-244-357
18	626	32.3	340	9	US-09-940-244-384
19					

#### ALIGNMENTS

RESULT 1  
US-09-805-311-2  
; Sequence 2, Application US/09805311  
; Patent No. US20010034886A1  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; TITLE OF INVENTION: Rad2/FEN-1 Orthologues and Uses  
; TITLE OF INVENTION: Thereof  
; FILE REFERENCE: 0961D  
; CURRENT APPLICATION NUMBER: US/09/805,311  
; CURRENT FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 09/426,557  
; PRIOR FILING DATE: 1999-10-22  
; PRIOR APPLICATION NUMBER: 60/112,332  
; PRIOR FILING DATE: 1998-12-15  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-805-311-2

Query Match	100.0%	Score 1939;	DB 10;	Length 379;
Best Local Similarity	100.0%	Pred. No. 2e+153;		
Matches 379;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGKGLTKLLADNAPKAMEKESYEGSKIAVDASMSYOFLLIVGRTGMEITLNEAGE	60	
Db	1	MGKGLTKLLADNAPKAMEKESYEGSKIAVDASMSYOFLLIVGRTGMEITLNEAGE	60	
Qy	61	VTSLGCMFNRTIRLEAGIKPVYFDGPPDKKQELAKRYSKRDDATKDLTFAVEVGD	120	
Db	61	VTSLGCMFNRTIRLEAGIKPVYFDGPPDKKQELAKRYSKRDDATKDLTFAVEVGD	120	
Qy	121	KDAIEKLSKRTYKVTQHNEDCKRLRLMGVVPVPEAPSEAECAALCINDKVFAYASED	180	
Db	121	KDAIEKLSKRTYKVTQHNEDCKRLRLMGVVPVPEAPSEAECAALCINDKVFAYASED	180	
Qy	181	MDSLTGAPRFLRLHMDSSKKIPVMEFDVAKVLELEITMDQFIDLCILCGDYCDSTK	240	
Db	181	MDSLTGAPRFLRLHMDSSKKIPVMEFDVAKVLELEITMDQFIDLCILCGDYCDSTK	240	

QY 241 GIGGOTALKLIHQSGTIESILENLNKRDIQIPEDWPYQEARRLFKEPNVTLDIPELKWTA 300  
 Db 241 GIGGOTALKLIHQSGTIESILENLNKRDIQIPEDWPYQEARRLFKEPNVTLDIPELKWTA 300  
 QY 301 PDEGLISFLVKGNGFNEQVTKAEIKSAKNSOGRLSEFFKPTATTAPLKRKETS 360  
 Db 301 PDEGLISFLVKGNGFNEQVTKAEIKSAKNSOGRLSEFFKPTATTAPLKRKETS 360  
 QY 361 DKTSKAAANKTKAGGKK 379  
 Db 361 DKTSKAAANKTKAGGKK 379

## RESULT 2

US-09-805-311-6  
 ; Sequence 6, Application US/09805311  
 ; Patent No. US20010034886A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mahajan, Pramod B.  
 ; TITLE OF INVENTION: Rad2/FEN-1 Orthologues and Uses  
 ; FILE REFERENCE: 0961D  
 ; CURRENT APPLICATION NUMBER: US/09/805,311  
 ; CURRENT FILING DATE: 2001-03-13  
 ; PRIOR APPLICATION NUMBER: 09/426,557  
 ; PRIOR FILING DATE: 1999-10-22  
 ; PRIOR APPLICATION NUMBER: 60/112,332  
 ; PRIOR FILING DATE: 1998-12-15  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 6  
 ; LENGTH: 379  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 US-09-805-311-6

Query Match 100.0%; Score 1939; DB 10; Length 379;  
 Best Local Similarity 100.0%; Pred. No. 2e-153;  
 Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGKGLTKLLADNAPKAMKEQKFESYFGKRTAVDASMSIYQFLIVVGRGTMTLTNEAGE 60  
 Db 1 MGKGLTKLLADNAPKAMKEQKFESYFGKRTAVDASMSIYQFLIVVGRGTMTLTNEAGE 60  
 QY 61 VTSHLQGFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120  
 Db 61 VTSHLQGFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120  
 QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVVPVVEAPSEAEACALCINDKVFVASED 180  
 Db 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVVPVVEAPSEAEACALCINDKVFVASED 180  
 QY 181 MDSLTFGAPRFLRLHMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCILCGCYCDSIK 240  
 Db 181 MDSLTFGAPRFLRLHMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCILCGCYCDSIK 240  
 QY 241 GIGGOTALKLIHQSGTIESILENLNKRDIQIPEDWPYQEARRLFKEPNVTLDIPELKWTA 300  
 Db 241 GIGGOTALKLIHQSGTIESILENLNKRDIQIPEDWPYQEARRLFKEPNVTLDIPELKWTA 300  
 QY 301 PDEGLISFLVKGNGFNEQVTKAEIKSAKNSOGRLSEFFKPTATTAPLKRKETS 360  
 Db 301 PDEGLISFLVKGNGFNEQVTKAEIKSAKNSOGRLSEFFKPTATTAPLKRKETS 360  
 QY 361 DKTSKAAANKTKAGGKK 379  
 Db 361 DKTSKAAANKTKAGGKK 379

## RESULT 3

US-09-805-311-4  
 ; Sequence 4, Application US/09805311

; Patent No. US20010034886A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mahajan, Pramod B.  
 ; TITLE OF INVENTION: Rad2/FEN-1 Orthologues and Uses  
 ; FILE REFERENCE: 0961D  
 ; CURRENT APPLICATION NUMBER: US/09/805,311  
 ; CURRENT FILING DATE: 2001-03-13  
 ; PRIOR APPLICATION NUMBER: 09/426,557  
 ; PRIOR FILING DATE: 1999-10-22  
 ; PRIOR APPLICATION NUMBER: 60/112,332  
 ; PRIOR FILING DATE: 1998-12-15  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 379  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 US-09-805-311-4

Query Match 99.7%; Score 1933; DB 10; Length 379;  
 Best Local Similarity 99.7%; Pred. No. 6.2e-153;  
 Matches 378; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MGKGLTKLLADNAPKAMKEQKFESYFGKRTAVDASMSIYQFLIVVGRGTMTLTNEAGE 60  
 Db 1 MGKGLTKLLADNAPKAMKEQKFESYFGKRTAVDASMSIYQFLIVVGRGTMTLTNEAGE 60  
 QY 61 VTSHLQGFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120  
 Db 61 VTSHLQGFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120  
 QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVVPVVEAPSEAEACALCINDKVFVASED 180  
 Db 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVVPVVEAPSEAEACALCINDKVFVASED 180  
 QY 181 MDSLTFGAPRFLRLHMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCILCGCYCDSIK 240  
 Db 181 MDSLTFGAPRFLRLHMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCILCGCYCDSIK 240  
 QY 241 GIGGOTALKLIHQSGTIESILENLNKRDIQIPEDWPYQEARRLFKEPNVTLDIPELKWTA 300  
 Db 241 GIGGOTALKLIHQSGTIESILENLNKRDIQIPEDWPYQEARRLFKEPNVTLDIPELKWTA 300  
 QY 301 PDEGLISFLVKGNGFNEQVTKAEIKSAKNSOGRLSEFFKPTATTAPLKRKETS 360  
 Db 301 PDEGLISFLVKGNGFNEQVTKAEIKSAKNSOGRLSEFFKPTATTAPLKRKETS 360  
 QY 361 DKTSKAAANKTKAGGKK 379  
 Db 361 DKTSKAAANKTKAGGKK 379

## RESULT 4

US-09-805-311-8  
 ; Sequence 8, Application US/09805311  
 ; Patent No. US20010034886A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mahajan, Pramod B.  
 ; TITLE OF INVENTION: Rad2/FEN-1 Orthologues and Uses  
 ; FILE REFERENCE: 0961D  
 ; CURRENT APPLICATION NUMBER: US/09/805,311  
 ; CURRENT FILING DATE: 2001-03-13  
 ; PRIOR APPLICATION NUMBER: 09/426,557  
 ; PRIOR FILING DATE: 1999-10-22  
 ; PRIOR APPLICATION NUMBER: 60/112,332  
 ; PRIOR FILING DATE: 1998-12-15  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 8  
 ; LENGTH: 379  
 ; TYPE: PRT

ORGANISM: Zea mays

US-09-805-311-8

Query Match 99.78; Score 1933; DB 10; Length 379;

Best Local Similarity 99.78; Pred No. 6.2e-15; Mismatches 0; Indels 0; Gaps 0;

Matches 378; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGKGLTKLLADNAPKAMKEQKPFESYGRKIAVDASMSIYQFLIVGRTGMETLTNEAGE 60  
Db 1 MGKGLTKLLADNAPKAMKEQKPFESYGRKIAVDASMSIYQFLIVGRTGMETLTNEAGE 60

Qy 61 VTSHLQGMNRTIRLEAGIKPVVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120  
Db 61 VTSHLQGMNRTIRLEAGIKPVVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

Qy 121 KDAIEKLSKRTVKVTRQNEDECKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180  
Db 121 KDAIEKLSKRTVKVTRQNEDECKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180

Qy 181 MDSLTFGAPRFLHLMPSKPKIPVMEFDVAKVLEELTMDQFIDLCILCGCDYCDISK 240  
Db 181 KDSLTFGAPRFLHLMPSKPKIPVMEFDVAKVLEELTMDQFIDLCILCGCDYCDISK 240

Qy 241 GIGGOTALKLIRQHSIESILENENKDRYQIPEDWPYQEARLKEPNTLDIPELKWTA 300  
Db 241 GIGGOTALKLIRQHSIESILENENKDRYQIPEDWPYQEARLKEPNTLDIPELKWTA 300

Qy 301 PDEGLISFLVKGNGFNEDRYTKAIEKISAKNKSOGRLSEFPKPTATTAPLKRKETS 360  
Db 301 PDEGLISFLVKGNGFNEDRYTKAIEKISAKNKSOGRLSEFPKPTATTAPLKRKETS 360

Qy 361 DKTSKAANKTKAGGKKK 379  
Db 361 DKTSKAANKTKAGGKKK 379

## RESULT 5

US-10-033-297-137  
; Sequence 137, Application US/10033297  
; Publication No. US20020187486A1

## GENERAL INFORMATION:

APPLICANT: Hall, Jeff G.  
; Lyamichiev, Victor I.  
; Mast, Andrea L.  
; Brow, Mary Ann D.

TITLE OF INVENTION: Detection of Nucleic Acids By Multiple  
; Sequential Invasive Cleavages

NUMBER OF SEQUENCES: 163

CORRESPONDENCE ADDRESS:

ADDRESS: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California

COUNTRY: United States Of America

ZIP: 94104

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/033,297

FILING DATE: 12-JAN-1997

CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/350,597

FILING DATE: 09-JUL-1999

APPLICATION NUMBER: US/08/823,516

FILING DATE: 24-MAR-1997

APPLICATION NUMBER: PCT/US97/01072

FILING DATE: 21-JAN-1997

APPLICATION NUMBER: US 08/759,038

FILING DATE: 02-DEC-1996

APPLICATION NUMBER: US 08/758,314  
; FILING DATE: 02-DEC-1996  
; APPLICATION NUMBER: US 08/756,386  
; FILING DATE: 29-NOV-1996  
; APPLICATION NUMBER: US 08/682,853  
; FILING DATE: 12-JUL-1996  
; APPLICATION NUMBER: US 08/599,491  
; FILING DATE: 24-JAN-1996

## ATTORNEY/AGENT INFORMATION:

NAME: Ingolla, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: FORS-02736

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

## INFORMATION FOR SEQ ID NO: 137:

## SEQUENCE CHARACTERISTICS:

LENGTH: 380 amino acids

TYPE: amino acid

STRANDEDNESS: No. US20020187486A1 Relevant

TOPOLOGY: No. US20020187486A1 Relevant

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 137:

US-10-033-297-137

Query Match 53.28; Score 1032; DB 9; Length 380;

Best Local Similarity 53.88; Pred No. 5.8e-78;

Matches 206; Conservative 69; Mismatches 98; Indels 10; Gaps 6;

Qy 1 MGKGLTKLLADNAPKAMKEQKPFESYGRKIAVDASMSIYQFLIVGRTGMETLTNEAGE 60  
Db 1 MGKGLTKLLADNAPKAMKEQKPFESYGRKIAVDASMSIYQFLIVGRTGMETLTNEAGE 60

Qy 61 VTSHLQGMNRTIRLEAGIKPVVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120  
Db 61 VTSHLQGMNRTIRLEAGIKPVVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

Qy 121 KDAIEKLSKRTVKVTRQNEDECKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180  
Db 121 KDAIEKLSKRTVKVTRQNEDECKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180

Qy 181 MDSLTFGAPRFLHLMPSKPKIPVMEFDVAKVLEELTMDQFIDLCILCGCDYCDISK 240  
Db 181 MDSLTFGAPRFLHLMPSKPKIPVMEFDVAKVLEELTMDQFIDLCILCGCDYCDISK 240

Qy 241 GIGGOTALKLIRQHSIESILENENKDRYQIPEDWPYQEARLKEPNTLDIPELKW 298  
Db 241 GIGGOTALKLIRQHSIESILENENKDRYQIPEDWPYQEARLKEPNTLDIPELKW 298

Qy 299 TAPDEGLISFLVKGNGFNEDRYTKAIEKISAKNKSOGRLSEFPKPTATTAPLKRKE 358  
Db 299 TAPDEGLISFLVKGNGFNEDRYTKAIEKISAKNKSOGRLSEFPKPTATTAPLKRKE 358

Qy 359 TSDK--TSKAANKTKAGGKKK 379  
Db 359 TSDK--TSKAANKTKAGGKKK 379

## RESULT 6

US-09-940-244-137

; Sequence 137, Application US/09940244

; Publication No. US2003004796A1

## GENERAL INFORMATION:

APPLICANT: Neiri, Bruce P.

APPLICANT: Hall, Jeff G.

APPLICANT: Lyamichiev, Victor

APPLICANT: Smith, Lloyd M.

TITLE OF INVENTION: Reactions on Dendrimers

FILE REFERENCE: FORS-06478

CURRENT APPLICATION NUMBER: US/09/940,244

CURRENT FILING DATE: 2002-05-06

NUMBER OF SEQ ID NOS: 422

SOFTWARE: Patent in version 3.1

SEQ ID NO 137  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-940-244-137

Query Match 53.2%; Score 1032; DB 9; Length 380;

Best Local Similarity 53.8%; Pred. No. 5.8e-78;  
Matches 206; Conservative 69; Mismatches 98; Indels 10; Gaps 6;

QY 1 MGIGLTKLADNAPKAMEQKESYGRKIAVDASMSIYQFLIVVGTGTMETITNEAGE 60  
DB 1 MGIGLTKLADNAPKAMEQKESYGRKIAVDASMSIYQFLIVVGTGTMETITNEAGE 60  
QY 61 VTSHQGMNRTIRLEAGIKPVYVFDGKPPDMKKOELAKRYSKRDDATKDLTAEVGD 120  
DB 60 TTSILMGMYRTIRMENGIRPVYVFDGKPPQLKSGELAKRSERRAEAKLOQAQAAGA 119  
QY 121 KDTEKLSKTKVTVRQHNEDCKRLRLMGVVPVVEAPSEAECAALCINDKVFVASED 180  
DB 120 EOEVEKTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAALVKAGKVAAATED 179  
QY 181 MDSITFCAPRFLRLMDPSSKKIPVMEFDVAKVLELELTMDQFIDLCLCGCDYCSIK 240  
DB 180 MDCITFGSPVIMRHLTASPAKRLQIEFHLRSLQELGNQOFVLDCLILGSDYCSIR 239  
QY 241 GIGQOTALKLIRHGSIESIENLNKORYQIPEDWPYQBARLFKEPNVTL--IPELKW 298  
DB 240 GIGPKRAVDLIQKHKSIEIEVRRDPNPKYVPENWNLKHAHOLEPEV-LDPESVELKW 298  
QY 299 TAPDEGLIFLVKNGFNEDRVTKAEIKSAKNKSSQGRLESFKEPTATTAPLKRKE 358  
DB 299 SEPNEELIKFMGCKEQSEIRSGVKRLSKRQSGTQGRDLDFKVTGSLSS-AKRKE 357  
QY 359 TSDK--TSAKAAANKTKAGKKK 379  
DB 358 PEPKSGSTKKA---KTGAAGKFK 377

## RESULT 7

US-09-805-311-10  
Sequence 10, Application US/09805311  
Patent No. US20010034886A1  
GENERAL INFORMATION:  
APPLICANT: Mahajan, Pramod B.  
TITLE OF INVENTION: Rad2/FEN-1 Orthologues and Uses  
FILE REFERENCE: 0961D  
CURRENT APPLICATION NUMBER: US/09/805,311  
PRIOR FILING DATE: 2001-03-13  
PRIOR APPLICATION NUMBER: 09/426,557  
PRIOR FILING DATE: 1999-10-22  
PRIOR APPLICATION NUMBER: 60/112,332  
PRIOR FILING DATE: 1998-12-15  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 10  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-805-311-10

Query Match 53.2%; Score 1032; DB 10; Length 380;

Best Local Similarity 53.8%; Pred. No. 5.8e-78;  
Matches 206; Conservative 69; Mismatches 98; Indels 10; Gaps 6;

QY 1 MGIGLTKLADNAPKAMEQKESYGRKIAVDASMSIYQFLIVVGTGTMETITNEAGE 60  
DB 1 MGIGLTKLADNAPKAMEQKESYGRKIAVDASMSIYQFLIVVGTGTMETITNEAGE 60  
QY 61 VTSHQGMNRTIRLEAGIKPVYVFDGKPPDMKKOELAKRYSKRDDATKDLTAEVGD 120  
DB 60 TTSILMGMYRTIRMENGIRPVYVFDGKPPQLKSGELAKRSERRAEAKLOQAQAAGA 119

QY 121 KDTEKLSKTKVTVRQHNEDCKRLRLMGVVPVVEAPSEAECAALCINDKVFVASED 180  
DB 120 EOEVEKTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAALVKAGKVAAATED 179  
QY 181 MDSITFCAPRFLRLMDPSSKKIPVMEFDVAKVLELELTMDQFIDLCLCGCDYCSIK 240  
DB 180 MDCITFGSPVIMRHLTASPAKRLQIEFHLRSLQELGNQOFVLDCLILGSDYCSIR 239  
QY 241 GIGQOTALKLIRHGSIESIENLNKORYQIPEDWPYQBARLFKEPNVTL--IPELKW 298  
DB 240 GIGPKRAVDLIQKHKSIEIEVRRDPNPKYVPENWNLKHAHOLEPEV-LDPESVELKW 298  
QY 299 TAPDEGLIFLVKNGFNEDRVTKAEIKSAKNKSSQGRLESFKEPTATTAPLKRKE 358  
DB 299 SEPNEELIKFMGCKEQSEIRSGVKRLSKRQSGTQGRDLDFKVTGSLSS-AKRKE 357  
QY 359 TSDK--TSAKAAANKTKAGKKK 379  
DB 358 PEPKSGSTKKA---KTGAAGKFK 377

RESULT 8  
US-10-033-297-138  
Sequence 138, Application US/10033297  
Publication No. US20020187486A1  
GENERAL INFORMATION:  
APPLICANT: Hall, Jeff G.  
Mast, Andrea L.  
Brow, Mary Ann D.  
Lyamichiev, Victor I.  
TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple Sequential Invasive Cleavages  
NUMBER OF SEQUENCES: 163  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/033,297  
FILING DATE: 12-NOV-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/350,597  
FILING DATE: 09-JUL-1999  
APPLICATION NUMBER: US/08/823,516  
FILING DATE: 24-MAR-1997  
APPLICATION NUMBER: PCT/US97/01072  
FILING DATE: 21-JAN-1997  
APPLICATION NUMBER: US 08/759,038  
FILING DATE: 02-DEC-1996  
APPLICATION NUMBER: US 08/758,314  
FILING DATE: 02-DEC-1996  
APPLICATION NUMBER: US 08/756,386  
FILING DATE: 29-NOV-1996  
APPLICATION NUMBER: US 08/682,853  
FILING DATE: 12-JUL-1996  
APPLICATION NUMBER: US 08/599,491  
FILING DATE: 24-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02736  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 138:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 amino acids  
TYPE: amino acid  
STRANDEDNESS: No. US20020187486A1 Relevant  
TOPOLOGY: No. US20020187486A1 Relevant  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 138:  
US-10-033-297-138

Query Match 52.3%; Score 1014; DB 9; Length 378;  
Best Local Similarity 53.2%; Pred. No. 1.8e-76;  
Matches 202; Conservative 69; Mismatches 101; Indels 8; Gaps 6;  
QY 1 MGIGLTKLADNAPKAMKEQFESYFGKRIADVADMSIYQFLIVVGRGTMTLTNEAGE 60  
DB 1 MGIGLTKLADNAPKAMKEQFESYFGKRIADVADMSIYQFLIVVGRGTMTLTNEAGE 60  
QY 61 VTSHLOGMFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVD 120  
DB 60 TTS-LMGMYRTIR-MENGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVD 117  
QY 121 KDAIEKLSKRTVKTROHNECKRLRLRLMGVVPVVEAPSEAECAALCINDKVFVASED 180  
DB 118 EEEVEKFTKRLVKVTKQHNDCKHLLSLMGIPYLDAPSEAECAALCINDKVFVASED 177  
QY 181 MDSLTGAPRFLRLHMDPSSKIPVMEFDVAKYLEELTMDQFIDLCILCGDCYCSIK 240  
DB 178 MDCITGSPVLMRLTASEAKLPQIEFHLKRVQLGELNQEQFVDCILILGSDYCESIR 237  
QY 241 GIGQATALKLIRHGSIESILENKNDRYQIPEDWPYQEARLKEPNVT-LDPELAKWT 299  
DB 238 GIGAKRAVDLIQKHKSIEEIVRLDPSKYVPYENWLHKEAQQLFLEPEVVDPESELKWS 297  
QY 300 APDEGLISFLVNDGNEFDRVTKAEIKKSAKNSQGRLESEFFKPTATTAPLARKKET 359  
DB 298 EPNEELVFMCGEKQFSEERIRSGVKRLSKRSQGSTQGRLLDDFFKVTGSLSS-AKRKEP 356  
QY 360 SDKTSKAAANKTKAGGKK 379  
DB 357 E---PKGPAKKAKTGGAGK 373

RESULT 9  
US-09-940-244-138  
Sequence 138, Application US/09940244  
Publication No. US20030044796A1  
GENERAL INFORMATION:  
APPLICANT: Neri, Bruce P.  
APPLICANT: Hall, Jeff G.  
APPLICANT: Lyamichev, Victor  
APPLICANT: Smith, Lloyd M.  
TITLE OF INVENTION: Reactions on Dendrimers  
FILE REFERENCE: FORS-06478  
CURRENT APPLICATION NUMBER: US/09/940,244  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 138  
LENGTH: 378  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-940-244-138

Query Match 52.3%; Score 1014; DB 9; Length 378;  
Best Local Similarity 53.2%; Pred. No. 1.8e-76;  
Matches 202; Conservative 69; Mismatches 101; Indels 8; Gaps 6;  
QY 1 MGIGLTKLADNAPKAMKEQFESYFGKRIADVADMSIYQFLIVVGRGTMTLTNEAGE 60  
DB 1 MGIGLTKLADNAPKAMKEQFESYFGKRIADVADMSIYQFLIVVGRGTMTLTNEAGE 59

QY 61 VTSHLOGMFNRTIRLLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVD 120  
DB 60 TTS-LMGMYRTIR-MENGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVD 117  
QY 121 KDAIEKLSKRTVKTROHNECKRLRLRLMGVVPVVEAPSEAECAALCINDKVFVASED 180  
DB 118 EEEVEKFTKRLVKVTKQHNDCKHLLSLMGIPYLDAPSEAECAALCINDKVFVASED 177  
QY 181 MDSLTGAPRFLRLHMDPSSKIPVMEFDVAKYLEELTMDQFIDLCILCGDCYCSIK 240  
DB 178 MDCITGSPVLMRLTASEAKLPQIEFHLKRVQLGELNQEQFVDCILILGSDYCESIR 237  
QY 241 GIGQATALKLIRHGSIESILENKNDRYQIPEDWPYQEARLKEPNVT-LDPELAKWT 299  
DB 238 GIGAKRAVDLIQKHKSIEEIVRLDPSKYVPYENWLHKEAQQLFLEPEVVDPESELKWS 297  
QY 300 APDEGLISFLVNDGNEFDRVTKAEIKKSAKNSQGRLESEFFKPTATTAPLARKKET 359  
DB 298 EPNEELVFMCGEKQFSEERIRSGVKRLSKRSQGSTQGRLLDDFFKVTGSLSS-AKRKEP 356  
QY 360 SDKTSKAAANKTKAGGKK 379  
DB 357 E---PKGPAKKAKTGGAGK 373

RESULT 10  
US-10-033-297-139  
Sequence 139, Application US/10033297  
Publication No. US20020187486A1  
GENERAL INFORMATION:  
APPLICANT: Hall, Jeff G.  
APPLICANT: Lyamichev, Victor I.  
APPLICANT: Mast, Andrea L.  
APPLICANT: Brown, Mary Ann D.  
TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
NUMBER OF SEQUENCES: 163  
CORRESPONDENCE ADDRESS:  
ADDRESS: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/033,297  
FILING DATE: 12-NOV-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/350,597  
FILING DATE: 09-Jul-1999  
APPLICATION NUMBER: US/08/823,516  
FILING DATE: 24-MAR-1997  
APPLICATION NUMBER: PCT/US97/01072  
FILING DATE: 21-JAN-1997  
APPLICATION NUMBER: US 08/759,038  
FILING DATE: 02-DEC-1996  
APPLICATION NUMBER: US 08/758,314  
FILING DATE: 02-DEC-1996  
APPLICATION NUMBER: US 08/756,386  
FILING DATE: 29-NOV-1996  
APPLICATION NUMBER: US 08/682,853  
FILING DATE: 12-JUL-1996  
APPLICATION NUMBER: US 08/599,491  
FILING DATE: 24-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: FORS-02736  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 139:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 382 amino acids  
TYPE: amino acid  
STRANDNESS: NO. US20020187486A1 Relevant  
TOPOLOGY: NO. US20020187486A1 Relevant  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 139:  
US-10-033-297-139

Query Match 48.4%; Score 938; DB 9; Length 382;  
Best Local Similarity 49.5%; Pred. No. 3.8e-70;  
Matches 190; Conservative 69; Mismatches 109; Indels 16; Gaps 6;

QY 1 MGIGLTKLLADNAPKAMKQKESYFGRIKAVDASMSIYQFLIYVGRGTMETITNEAGE 60  
DB 1 MGIGLTKLLADNAPKAMKQKESYFGRIKAVDASMSIYQFLIYVGRGTMETITNEAGE 60  
QY 61 VTSHLQGMFNRTIRLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120  
DB 61 TTSHLMGMFYRTLRMDINGIKPCYVFDGKPPDLKSHLTKRSSRVETEKKLAETTELE 120  
QY 121 KDAIEKLSKRTVKVTRHNEDECKRLRLMGVVPVVEAPSEAEACALCINDKVFAYASED 180  
DB 121 K--MQERRLVKVSKEHNEEAQKLLGLMGIPYIAPTAEAAQCAELAKKGVYAAASED 177  
QY 181 MDSLTGAPRFLRHLMPPSSKKIPVMEFDVAKVLELELTMDQFDICILGCDYCDISIK 240  
DB 181 MDTLCYRTPFLRLHLTFFSEAKKEPIHEIDTLVLRGLDITIEQFVLDLCIMLCCDYCESIR 237  
QY 241 GIGGOTALKLIHQHGSIESILENL-----NKDRYQIPEDWPYQEARLKFEPNVTLDIPE 295  
DB 241 GIGGOTALKLIHQHGSIESILENL-----NKDRYQIPEDWPYQEARLKFEPNVTLDIPE 295  
QY 296 --LKTAPDEGLISFLVKGNGFNEEDRVTKAIEKIKSAKNKSSQGRLESFFKPTATT--- 350  
DB 296 --LKTAPDEGLISFLVKGNGFNEEDRVTKAIEKIKSAKNKSSQGRLESFFKPTATT--- 350  
QY 351 -SAPLKRKETSOKTSAKAAANKTK 373  
DB 357 LAAAKRAQENKLNK-NKNKVTK 379

RESULT 11  
US-09-940-244-139  
Sequence 139, Application US/09940244  
Publication No. US20030044796A1  
GENERAL INFORMATION:  
APPLICANT: Neri, Bruce P.  
APPLICANT: Hall, Jeff G.  
APPLICANT: Lyamichiev, Victor  
APPLICANT: Smith, Lloyd M.  
TITLE OF INVENTION: Reactions on Dendrimers  
FILE REFERENCE: FORS-06478  
CURRENT FILING DATE: 2002-05-06  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 139  
LENGTH: 382  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-09-940-244-139

Query Match 48.4%; Score 938; DB 9; Length 382;  
Best Local Similarity 49.5%; Pred. No. 3.8e-70;  
Matches 190; Conservative 69; Mismatches 109; Indels 16; Gaps 6;  
QY 1 MGIGLTKLLADNAPKAMKQKESYFGRIKAVDASMSIYQFLIYVGRGTMETITNEAGE 60

DB 1 MGIGLTKLLADNAPKAMKQKESYFGRIKAVDASMSIYQFLIYVGRGTMETITNEAGE 60  
QY 61 VTSHLQGMFNRTIRLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120  
DB 61 TTSHLMGMFYRTLRMDINGIKPCYVFDGKPPDLKSHLTKRSSRVETEKKLAETTELE 120  
QY 121 KDAIEKLSKRTVKVTRHNEDECKRLRLMGVVPVVEAPSEAEACALCINDKVFAYASED 180  
DB 121 K--MQERRLVKVSKEHNEEAQKLLGLMGIPYIAPTAEAAQCAELAKKGVYAAASED 177  
QY 181 MDSLTGAPRFLRHLMPPSSKKIPVMEFDVAKVLELELTMDQFDICILGCDYCDISIK 240  
DB 181 MDTLCYRTPFLRLHLTFFSEAKKEPIHEIDTLVLRGLDITIEQFVLDLCIMLCCDYCESIR 237  
QY 241 GIGGOTALKLIHQHGSIESILENL-----NKDRYQIPEDWPYQEARLKFEPNVTLDIPE 295  
DB 241 GIGGOTALKLIHQHGSIESILENL-----NKDRYQIPEDWPYQEARLKFEPNVTLDIPE 295  
QY 296 --LKTAPDEGLISFLVKGNGFNEEDRVTKAIEKIKSAKNKSSQGRLESFFKPTATT--- 350  
DB 296 --LKTAPDEGLISFLVKGNGFNEEDRVTKAIEKIKSAKNKSSQGRLESFFKPTATT--- 350  
QY 351 -SAPLKRKETSOKTSAKAAANKTK 373  
DB 357 LAAAKRAQENKLNK-NKNKVTK 379

RESULT 12  
US-10-033-297-79  
Sequence 79, Application US/10033297  
Publication No. US20020187486A1  
GENERAL INFORMATION:  
APPLICANT: Hall, Jeff G.  
APPLICANT: Lyamichiev, Victor I.  
APPLICANT: Mast, Andrea L.  
APPLICANT: Brow, Mary Ann D.  
TITLE OF INVENTION: Detection of Nucleic Acids By Multiple  
NUMBER OF SEQUENCES: 163  
CORRESPONDENCE ADDRESS: 163  
ADDRESS: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/033,297  
FILING DATE: 12-No. US20020187486A1-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/350,597  
FILING DATE: 09-Jul-1999  
APPLICATION NUMBER: US/08/823,516  
FILING DATE: 24-MAR-1997  
APPLICATION NUMBER: PCT/US97/01072  
FILING DATE: 21-JAN-1997  
APPLICATION NUMBER: US 08/759,038  
FILING DATE: 02-DEC-1996  
APPLICATION NUMBER: US 08/758,314  
FILING DATE: 02-DEC-1996  
APPLICATION NUMBER: US 08/756,386  
FILING DATE: 29-NOV-1996  
APPLICATION NUMBER: US 08/682,853  
FILING DATE: 12-JUL-1996  
APPLICATION NUMBER: US 08/599,491  
FILING DATE: 24-JAN-1996



FILE REFERENCE: FORS-06478  
CURRENT APPLICATION NUMBER: US/09/940,244  
CURRENT FILING DATE: 2002-05-06  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 79  
LENGTH: 340  
TYPE: PRT  
ORGANISM: Pyrococcus furiosus  
US-09-940-244-79

Query Match 33.4%; Score 648.5; DB 9; Length 340;

Best Local Similarity 41.3%; Pred. No. 4e-46;  
Matches 142; Conservative 62; Mismatches 107; Indels 33; Gaps 6;

QY 19 KEQFESYFGKRIADVDSMSIQFLIVVGRFTMETLTNEAGEVTSHLQGMFNRTIRLEA 78  
DB 12 KEIELENYGKKIADALNAIQFLSTIRQDGTPLMDSKGRITSHLSGLFYRTINLMEA 71  
QY 79 GIKPVYVFDGKPPDMKKOELAKRYSKRDDATKDLTEAVEVGDKDAIEKLSKRTVKVTRQH 138  
DB 72 GIKPVYVFDGEPPEKKKEKREAREAEKREAREALEKGBIEARKYAQRATRVNEML 131  
QY 139 NEDCKRLRLMGVVPVAPSAEACALCINDKVFVAVASEMDSLTFCGAPRFLRLHMDP 198  
DB 132 IEDAKKLELMGPIVQAPSEGEAAQAAVMAAGSVYASASQDYDLSLLFGAPRLVRLNLTIT 191  
QY 199 SSKKIP-----VMEFDVAKVLELELTMDQFIDLCILCGCDY-CDSIKGIGG 244  
DB 192 GKRLPGKNVYVVEIKPELIIIE-----EVLKELKLTREKLELAILVGTDPNPGGKIGIGL 247  
QY 245 QTALKLIRQHGSIESTILENLKND---RYQIPEDWPYQEARLFKEPNVTLDIPELKWTPAP 301  
DB 248 KKALEIVRH-----SKDPLAKFQKQSDVDLYAIKEFFLNPPVT-DNYNLVWRDP 295  
QY 302 DEGLISFLYKNGFNEDRVTKAIEKISAKNKSQGRLESFFK 345  
DB 296 DEEGILAFCLDHDHDFSEERVKNGLERLAKKAIKSGKSTLESNFK 339

## RESULT 15

US-09-940-244-136  
Sequence 136, Application US/09940244  
Publication No. US20030044796A1  
GENERAL INFORMATION:  
APPLICANT: Neri, Bruce P.  
APPLICANT: Hall, Jeff G.  
APPLICANT: Lyamichev, Victor  
APPLICANT: Smith, Lloyd M.  
TITLE OF INVENTION: Reactions on Dendrimers  
FILE REFERENCE: FORS-06478  
CURRENT APPLICATION NUMBER: US/09/940,244  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 136  
LENGTH: 340  
TYPE: PRT  
ORGANISM: Pyrococcus furiosus  
US-09-940-244-136

Query Match 33.4%; Score 648.5; DB 9; Length 340;

Best Local Similarity 41.3%; Pred. No. 4e-46;  
Matches 142; Conservative 62; Mismatches 107; Indels 33; Gaps 6;

QY 19 KEQFESYFGKRIADVDSMSIQFLIVVGRFTMETLTNEAGEVTSHLQGMFNRTIRLEA 78  
DB 12 KEIELENYGKKIADALNAIQFLSTIRQDGTPLMDSKGRITSHLSGLFYRTINLMEA 71  
QY 79 GIKPVYVFDGKPPDMKKOELAKRYSKRDDATKDLTEAVEVGDKDAIEKLSKRTVKVTRQH 138  
DB 72 GIKPVYVFDGEPPEKKKEKREAREAEKREAREALEKGBIEARKYAQRATRVNEML 131

QY 139 NEDCKRLRLMGVVPVAPSAEACALCINDKVFVAVASEMDSLTFCGAPRFLRLHMDP 198  
DB 132 IEDAKKLELMGPIVQAPSEGEAAQAAVMAAGSVYASASQDYDLSLLFGAPRLVRLNLTIT 191  
QY 199 SSKKIP-----VMEFDVAKVLELELTMDQFIDLCILCGCDY-CDSIKGIGG 244  
DB 192 GKRLPGKNVYVVEIKPELIIIE-----EVLKELKLTREKLELAILVGTDPNPGGKIGIGL 247  
QY 245 QTALKLIRQHGSIESTILENLKND---RYQIPEDWPYQEARLFKEPNVTLDIPELKWTPAP 301  
DB 248 KKALEIVRH-----SKDPLAKFQKQSDVDLYAIKEFFLNPPVT-DNYNLVWRDP 295  
QY 302 DEGLISFLYKNGFNEDRVTKAIEKISAKNKSQGRLESFFK 345  
DB 296 DEEGILAFCLDHDHDFSEERVKNGLERLAKKAIKSGKSTLESNFK 339

Search completed: May 8, 2003, 21:45:42  
Job time : 59 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:36:12 ; Search time 44 Seconds  
(without alignments)  
828.067 Million cell updates/sec

Title: US-09-805-311-2

Perfect score: 1939

Sequence: 1 MGIGKLTLLADNAPKAMKE.....SDKTSKAAANKTKAGKKK 379

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR73:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1313.5	67.7	362	2 T01198	endonuclease homol
2	1032	53.2	380	2 A56531	DNA structure-spec
3	1023.5	52.8	380	2 A56054	DNA repair endonuc
4	1014	52.3	378	2 A53730	flap endonuclease
5	998.5	51.5	385	2 T13692	hypothetical prote
6	938	48.4	382	2 S2287	hypothetical prote
7	648.5	33.4	340	2 T46893	endo/exonuclease f
8	636	32.8	343	2 E75117	DNA repair protein
9	632.5	32.6	343	2 A71015	probable 5' nuclea
10	560.5	28.9	326	2 C84480	DNA repair protein
11	542	28.0	328	2 C59085	DNA repair protein
12	538.5	27.8	336	2 H69282	DNA repair protein
13	514.5	26.5	401	2 H72765	probable flap endo
14	506	26.1	302	2 F90158	DNA repair endo/ex
15	369	19.0	327	2 F94290	DNA repair protein
16	336	17.3	1516	2 E71619	RAD2 endonuclease
17	267.5	13.8	317	2 E86419	probable exonuclea
18	260.5	13.4	571	2 T43288	probable exodeoxyr
19	245.5	12.7	496	2 S83300	probable excision
20	235	12.1	497	2 F86150	F22M8.2 protein -
21	232.5	12.0	1031	2 A29839	RAD2 protein - yea
22	230.5	11.4	1186	2 S35993	DNA repair protein
23	220	11.3	1170	2 A57650	repair protein xpg
24	218	11.2	702	2 S62169	DHS1 protein - yea
25	206	10.6	1113	2 S30301	excision repair pr
26	204.5	10.5	430	2 T40382	DIM7 protein - yea
27	204.5	10.5	1112	2 T40382	dna repair protein
28	202	10.4	1185	2 T58009	gene ERCC5 protein
29	200	10.3	1196	2 S35994	DNA repair protein

30 197 10.2 1287 2 T22235 hypothetical prote  
31 179.5 9.3 829 2 T32744 hypothetical prote  
32 176.5 9.1 986 2 S77056 DNA-directed DNA p  
33 173 8.9 832 2 A33530 DNA-directed DNA p  
34 172.5 8.9 434 2 T34350 hypothetical prote  
35 171 8.8 831 2 JX0359 DNA-directed DNA p  
36 166 8.6 289 2 D70440 DNA polymerase I, 5'-3' exonuclease  
37 162.5 8.4 871 2 E97035 protein T10022.7  
38 158.5 8.2 290 2 A11309 DNA-directed DNA p  
39 157.5 8.1 831 2 S26675 DNA-directed DNA p  
40 156.5 8.1 877 2 H97875 DNA-directed DNA p  
41 156.5 8.1 877 2 H97875 DNA polymerase I  
42 154.5 8.0 877 2 E95003 hypothetical prote  
43 154.5 8.0 337 2 T49286 hypothetical prote  
44 151.5 7.8 759 2 S50544 hypothetical prote  
45 151.5 7.8 759 2 S50544 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

T01198

endonuclease homolog F21E10.3 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 22-Oct-1999

C:Accession: T01198

R:Davidson, S.; Rohlfing, T.; David, M.; O'Brian, D.

submitted to the EMBL Data Library, April 1998

A:Description: The sequence of A. thaliana F21E10.

A:Reference number: Z14258

A:Accession: T01198

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-362 <DAV>

A:Cross-references: EMBL:AF058914; NID:g3047074; PID:g3047083; PIDN:AAC13596.1; GSPDB

A:Experimental source: cultivar Columbia

C:Genetics:

A:Gene: ATSP:F21E10.3

A:Map position: 5

A:Introns: 4/3; 51/2; 71/1; 96/2; 116/2; 134/2; 150/3; 166/3; 188/3; 205/3; 245/3; 27

C:Superfamily: conserved hypothetical protein YKL113c

C:Keywords: DNA repair; endonuclease

Query Match 67.7%; Score 1313.5; DB 2; Length 362;

Best Local Similarity 72.1%; Pred. No. 1.5e-80;

Matches 259; Conservative 30; Mismatches 33; Indels 37; Gaps 3;

QY 1 MGIGKLTLLADNAPKAMKEOKFESYFGRIADVDSMSYQPLIVVGRGTMETLTNEAGE 60

Db 1 MGIGKLTLLADNAPKAMKEOKFESYFGRIADVDSMSYQPLIVVGRGTMETLTNEAGE 60

QY 61 VTS-----HLOGMFRNTRIRLEAGIKPVYVF 86

Db 60 KCSLMKLVKSLWIFLCLVLYVFFVYVYVMMFNRHLOGMFRNTRIRLEAGIKPVYVF 119

QY 87 DCKPPDMKKQLAKRYSKRDATDKLTAEVGVGDKDIKLSKRTVKVYTRHNDCKRL 146

Db 120 DCKPELLARQELAKRYSKRADATDLTGAEGNKEDIEKYSKRTVKVYTRHNDCKRL 179

QY 147 RLMGVVVEATSEAEACALCNDKVFVASEDMDSITFGAPRFLHMDPSSKKIPVM 206

Db 180 RLMGVVVEATSEAEACALCNDKVFVASEDMDSITFGAPRFLHMDPSSKKIPVM 239

QY 207 EFDVAKVLEELPMDFIDLCILCGDCYCDSTIGIGGTALKLRQHGSTESILENLK 266

Db 240 EPEVAKILEELQMTDFIDLCILCGDCYCDSTIGIGGTALKLRQHGSTETILENLK 299

QY 267 DRYQTPEDWPQEARLFKEPNVTLPIEL--KWTAPDEEGLISFLYKDNCFNDRVK 323

Db 300 ERYQTPEDWPQEARLFKEPNVTLPIEL--KWTAPDEEGLISFLYKDNCFNDRVK 358

## RESULT 2

A56531

DNA structure-specific endonuclease FEN1 [validated] - human

N:Alternate names: flap endonuclease 1; rad2 protein homolog

C:Species: Homo sapiens (man)

C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 21-Jul-2000

A:Accession: A56531; B56054

R:Hiraoka, L.R.; Harrington, J.J.; Gerhard, D.S.; Lieber, M.R.; Hsieh, C.L.

Genomics 25, 220-225, 1995

A:Title: Sequence of human FEN-1, a structure-specific endonuclease and chromosomal local

A:Reference number: A56531; MUID:95293376; PMID:7774922

A:Accession: A56531

A&gt;Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-380 &lt;HIR&gt;

A:CROSS-references: GB:L37374; NID:9642089; PIDN:AAA91331.1; PID:G642090

R:Murray, J.M.; Tavassoli, M.; Al-Harithy, R.; Sheldrick, K.S.; Lehmann, A.R.; Carr, A.M.

Mol. Cell. Biol. 14, 4878-4888, 1994

A:Title: Structural and functional conservation of the human homolog of the Schizosaccha

A:Reference number: A56054; MUID:94277093; PMID:8007985

A:Accession: B56054

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-380 &lt;MUR&gt;

A:CROSS-references: GB:S70527

C:Genetics:

A:Gene: GDB:FEN1

A:CROSS-references: GDB:451178; OMIM:600393

A:Map position: 11q12-11q12

C:Function:

A:Description: cleaves substrates with unannealed 5'-tails [validated, MUID:20209428]

A:Note: needed during Okazaki fragment processing and long patch base excision repair,

C:Superfamily: conserved hypothetical protein YKL113C

C:Keywords: DNA repair; endonuclease

Query Match 53.2%; Score 1032; DB 2; Length 380;  
 Best Local Similarity 53.8%; Pred. No. 1,1e-61;  
 Matches 206; Conservative 69; Mismatches 98; Indels 10; Gaps 6;

QY 1 MGKGLTKLLADNAPKAMKQKESYFGKRIADVADSMSTYQFLIVVGRGTMETLTNEAGE 60

DB 1 MGIOGLAKLADVAPSARENDIKSYFGKRIADVADSMSTYQFLIVVGRGTMETLTNEAGE 59

QY 61 VTSHLQGMFNRTIRLLAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

DB 60 TTSHLMGMFYRTIRMENGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 119

QY 121 KDAIEKLSKRTVKYTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180

DB 120 EDEVERKTRVLKVTQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 179

QY 181 MDSLTFGAPRFLRLHMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCILGCDYCDISIK 240

DB 180 MDCLTFGAPVRLHMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCILGCDYCDISIR 239

QY 241 GIGGOTALKLIRQHSISILENLKDRYQIPEDMPYQEARLFKEPNVTLD--IPELKW 298

DB 240 GIGPRAVDLLQKHSIEIVRLDPNKYPVPMNLHKEAHLFEPEV-LDPESVELKW 298

QY 299 TAPDEGLISFLVKGDNFNRVTKALEIKSAKNSSQGRLESFFKPTATTAPLAKKE 358

DB 299 SEPNEELIKPWCCEKOFSEERTISGVKRLSKRSQSGTQGLRDLDFKVTGSLSS-AKKE 357

QY 359 TSDK--TSAANKTKTAGGKK 379

DB 358 PEPKGTSTKKA---KTGAAGKFK 377

## RESULT 3

A56054

DNA repair endonuclease rad2 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 21-Jul-2000

C:Accession: A56054; T45210; T38725  
 R:Murray, J.M.; Tavassoli, M.; Al-Harithy, R.; Sheldrick, K.S.; Lehmann, A.R.; Carr, A.M.

Mol. Cell. Biol. 14, 4878-4888, 1994

A:Title: Structural and functional conservation of the human homolog of the Schizosac

A:Reference number: A56054; MUID:94277093; PMID:8007985

A:Accession: A56054

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-380 &lt;MUR&gt;

A:CROSS-references: GB:S70605

A:Accession: T45210

A&gt;Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-380 &lt;MUR&gt;

A:CROSS-references: EMBL:X77041; NID:92909339; PIDN:CAB36991.1; PID:94456854

A:Experimental source: strain 972h(-)

R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, September 1997

A:Reference number: Z21797

A:Accession: T38725

A&gt;Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-380 &lt;GEN&gt;

A:CROSS-references: EMBL:Z99167; PIDN:CAB16282.1; GSPDB:GNO0066; SPDB:SPAC3G6.06C

A:Experimental source: strain 972h; cosmid C3G6

C:Genetics:

A:Gene: rad2

A:Map position: 1

A:Introns: 5/1; 35/2

C:Function:

A:Description: DNA repair

C:Superfamily: conserved hypothetical protein YKL113C

Query Match 52.8%; Score 1023.5; DB 2; Length 380;  
 Best Local Similarity 50.7%; Pred. No. 3.9e-61;  
 Matches 194; Conservative 73; Mismatches 107; Indels 9; Gaps 3;

QY 1 MGKGLTKLLADNAPKAMKQKESYFGKRIADVADSMSTYQFLIVVGRGTMETLTNEAGE 60

DB 1 MGKGLTKLLADNAPKAMKQKESYFGKRIADVADSMSTYQFLIVVGRGTMETLTNEAGE 60

QY 61 VTSHLQGMFNRTIRLLAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

DB 61 TTSHLMGMFYRTIRLVNDNGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120

QY 121 KDAIEKLSKRTVKYTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180

DB 121 AEMVDFAKRTVKYTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180

QY 181 MDSLTFGAPRFLRLHMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCILGCDYCDISIK 240

DB 181 MDCLTFGAPVRLHMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCILGCDYCDISIR 240

QY 241 GIGGOTALKLIRQHSISILENLKDRYQIPEDMPYQEARLFKEPNVTLDIP----EL 296

DB 241 GIGPRAVDLLQKHSIEIVRLDPNKYPVPMNLHKEAHLFEPEV-LDPESVELKW 297

QY 297 TAPDEGLISFLVKGDNFNRVTKALEIKSAKNSSQGRLESFFKPTATTAPLAKR 356

DB 298 KWKSPDAGDIQFLVKEGFDNRVTKALEIKSAKNSSQGRLESFFKPTATTAPLAKR 355

QY 357 KETSDEKTSKAAANKTKTAGGKK 379

DB 356 VDTKSKGSAKRKRSNKGESKK 378

## RESULT 4

A53730

flap endonuclease 1 - mouse

N:Alternate names: FEN-1

C:Species: Mus musculus (house mouse)

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 24-Sep-1999

C:Accession: A53730

R:Harrington, J.J.; Lieber, M.R.  
 Genes Dev 8, 1344-1355, 1994  
 A:Title: Functional domains within FEN-1 and RAD2 define a family of structure-specific  
 A:Reference number: A53730; MUID:95011546; PMID:7926735  
 A:Accession: A53730  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-378 <HAR>  
 A:Cross-references: GB:L26320; NID:g499648; PIDN:AAC37664.1; PID:g499649  
 C:Superfamily: conserved hypothetical protein YKL113c  
 C:Keywords: DNA repair; endonuclease

Query Match 52.3%; Score 1014; DB 2; Length 378;  
 Best Local Similarity 53.2%; Pred. No. 1.7e-50;  
 Matches 202; Conservative 69; Mismatches 101; Indels 8; Gaps 6;  
 QY 1 MGIGLTKLLADNAPKAMKEQKESYFGKRTAVDASMSIQFLIVVGRGTMTLTNEAGE 60  
 DB 1 MGIGLTKLLADNAPKAMKEQKESYFGKRTAVDASMSIQFLIVVGRGTMTLTNEAGE 59  
 QY 61 VTSHLOGMENRTIRLEAGIKPVYVDFGKPPDMKKOELAKRYSKRDDATFDTLTAEEVGD 120  
 DB 60 TTS-LMGMEFRTIR-WENGIRKPVYVDFGKPPDLKSGELAKRYSKRAEAEKQOQAEAGM 117  
 QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180  
 DB 118 EEEVERKTRKLVKVKQHNDECKHLLSLMGIPYLDAPSEAEACALAKAGKVVYAAATED 177  
 QY 181 MDSLTGAPRFLRLHMDPSSKKIPVMEFDNAKVLLEELTMDQFIDLCILCGDCYCSIK 240  
 DB 178 MDCLTGSPVLMRLHMTASEAKKLTQEPHLSRVLQELGLNQEFVDCILGLGSDYCSIR 237  
 QY 241 GIGGOTALKLIRHOGSIESILENLKDRYQIPEDWPYQEARRLFKEPNVT-LDIPELKWT 299  
 DB 238 GIGAKRAVDLQKHKSIEIRVRLDPSKYPVPENLWKEAQLFLPEVVDPESEVELKWS 297  
 QY 300 APDEGLISFLVKNDFNEDRVTKAIEKIKSAKNSQGRLESFFK-----PTATTSAPL 354  
 DB 298 EPNEELVPMGCEKQFSEIRSRVSKRSQSGTQGLDDOFFKVTGSLSS-AKRKEP 356  
 QY 360 SDKTSKAAANKTKAGGKKK 379  
 DB 357 E---PKGPAKKKAKTGAGK 373

RESULT 5  
 T13692  
 Hypothetical protein EG0003.3 - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C:Accession: T13692  
 R:Murphy, L.; Harris, D.; Barrell, B.  
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
 A:Reference number: Z17699  
 A:Accession: T13692  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-385 <MUR>  
 A:Cross-references: EMBL:AL031863; PIDN:CAA21320.1  
 C:Genetics:  
 A:Introns: 28/1  
 A:Note: EG:EG0003.3  
 C:Superfamily: conserved hypothetical protein YKL113c

Query Match 51.5%; Score 998.5; DB 2; Length 385;  
 Best Local Similarity 52.9%; Pred. No. 1.9e-59;  
 Matches 202; Conservative 65; Mismatches 104; Indels 11; Gaps 5;  
 QY 1 MGIGLTKLLADNAPKAMKEQKESYFGKRTAVDASMSIQFLIVVGRGTMTLTNEAGE 60  
 DB 1 MGIGLTKLLADNAPKAMKEQKESYFGKRTAVDASMSIQFLIVVGRGTMTLTNEAGE 59

QY 61 VTSHLOGMENRTIRLEAGIKPVYVDFGKPPDMKKOELAKRYSKRDDATFDTLTAEEVGD 120  
 DB 60 PTSHLMGMEYRTIRLLDNGIKPVYVDFGKPPDLKSGELAKRYSKRAEAEKALKAAATDAGD 119  
 QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180  
 DB 120 DAGIEKFNRLRVTRVTHAKAEKELLTLMGVYVDAPCEAEACALAKAGKVVATATED 179  
 QY 181 MDSLTGAPRFLRLHMDPSSKKIPVMEFDNAKVLLEELTMDQFIDLCILCGDCYCSIK 240  
 DB 180 MDALTGSKLLRYLTYSEARKMPKVESYDKLLEGLAINNREFIDLCILCGDCYCSIK 239  
 QY 241 GIGGOTALKLIRHOGSIESILENLKDRYQIPEDWPYQEARRLFKEPNVT-LDIPELKWT 299  
 DB 240 GIGKRAIELINTYRDIEFILNDSSKTYVPENWTKVARELFIEPEVADADSIDLKVV 299  
 QY 300 APDEGLISFLVKNDFNEDRVTKAIEKIKSAKNSQGRLESFFK-----PTATTSAPL 354  
 DB 300 EPDEGLVFLCGDRQFNEERVNGAKMLKMSKQAQTVRLDSFFKTLSTPNATNAA-- 357  
 QY 355 KRKETSCKTSAKAAANKTKAGG 376  
 DB 358 KRK--AEAKKSANKKAKTSG 377

RESULT 6  
 S22267  
 Hypothetical protein YKL113c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein YKL510; RAD2 protein homolog  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 24-Sep-1999  
 C:Accession: S22267; S37941  
 R:Jacquier, A.; Legrain, P.; Dujon, B.  
 A:Title: Sequence of a 10.7 kb segment of yeast chromosome XI identifies the APN1 an  
 s.  
 A:Reference number: S22267; MUID:92221689; PMID:1561835  
 A:Accession: S22267  
 A:Molecule type: DNA  
 A:Residues: 1-382 <JAC>  
 A:Cross-references: GB:S93804; NID:g248391; PIDN:AAB21998.1; PID:g248392  
 A:Experimental source: strain S288C  
 R:Jacquier, A.; Legrain, P.; Collea, L.; Richard, G.F.; Thierry, A.; Dujon, B.  
 A:Submitted to the Protein Sequence Database, March 1994  
 A:Reference number: S37938  
 A:Accession: S37941  
 A:Molecule type: DNA  
 A:Residues: 1-382 <JAC>  
 A:Cross-references: EMBL:Z28113; NID:g486189; PIDN:CAA81953.1; PID:g486190; MIPS:YKL1  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Gene: SGD:RAD27  
 A:Cross-references: SGD:S0001596; MIPS:YKL113c  
 A:Map position: 11L  
 C:Superfamily: conserved hypothetical protein YKL113c  
 C:Keywords: nucleus

Query Match 48.4%; Score 938; DB 2; Length 382;  
 Best Local Similarity 49.5%; Pred. No. 2.1e-55;  
 Matches 190; Conservative 69; Mismatches 109; Indels 16; Gaps 6;  
 QY 1 MGIGLTKLLADNAPKAMKEQKESYFGKRTAVDASMSIQFLIVVGRGTMTLTNEAGE 60  
 DB 1 MGIGLTKLLADNAPKAMKEQKESYFGKRTAVDASMSIQFLIVVGRGTMTLTNEAGE 60  
 QY 61 VTSHLOGMENRTIRLEAGIKPVYVDFGKPPDMKKOELAKRYSKRDDATFDTLTAEEVGD 120  
 DB 61 TTSHLMGMEYRTIRLLDNGIKPVYVDFGKPPDLKSGELAKRYSKRAEAEKALKAAATDAGD 120  
 QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180  
 DB 121 K--MKQERLVKVSKEHNEEAQKLLGIMGIPYIATPEAEACALAKAGKVVAAASED 177

Query Match 48.4%; Score 938; DB 2; Length 382;  
 Best Local Similarity 49.5%; Pred. No. 2.1e-55;  
 Matches 190; Conservative 69; Mismatches 109; Indels 16; Gaps 6;  
 QY 1 MGIGLTKLLADNAPKAMKEQKESYFGKRTAVDASMSIQFLIVVGRGTMTLTNEAGE 60  
 DB 1 MGIGLTKLLADNAPKAMKEQKESYFGKRTAVDASMSIQFLIVVGRGTMTLTNEAGE 60  
 QY 61 VTSHLOGMENRTIRLEAGIKPVYVDFGKPPDMKKOELAKRYSKRDDATFDTLTAEEVGD 120  
 DB 61 TTSHLMGMEYRTIRLLDNGIKPVYVDFGKPPDLKSGELAKRYSKRAEAEKALKAAATDAGD 120  
 QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180  
 DB 121 K--MKQERLVKVSKEHNEEAQKLLGIMGIPYIATPEAEACALAKAGKVVAAASED 177





A:Cross-references: GB:AE001087; GB:AE000782; NID:g2689410; PIDN:AA890967.1; PID:g265037  
C:Superfamily: conserved hypothetical protein YKL113c

Query Match 27.8%; Score 538.5; DB 2; Length 336;  
Best Local Similarity 36.7%; Pred. No. 9.3e-29;  
Matches 125; Conservative 67; Mismatches 118; Indels 31; Gaps 7;

QY 19 KEOKPESYGRKIAVDASMSIYQELIVVGRGTMETLTNEAGEVTSKLOGMENRTIRLLEA 78  
DB 12 EEVELEYSGKKIADVAENTLYQISIRQDPGTPLKDSQGRITSHLSILYRVSNWVEV 71

QY 79 GKPVVYVFGKPPDMKKQELAKRYSKRDATKDLTEAVEVGDKDAIEKLSKRTVKVTRQH 138  
DB 72 GIRPVVFGPEPFKAEIERKKRAEAEEMWIAALQAGKDA-KKYAAQAGRVDYI 130

QY 139 NEDCKRLRLMGVVPVVEAPSAEAEACALCINDKVFVASEDMDSITFGAPFLRLMDP 198  
DB 131 VDSAKTLLSYMGIPFVDPAPSEGAQAAYMAAGDVETGSDYDSLLFGSPRLARNLAIT 190

QY 199 SSKKIP-----VMEFDVAKVLELELTMDQFIDLCILGCDYCSIGIGGQ 245  
DB 191 GKRKLPGKNVYVDVPEIILLESN-----LKRGLTREQLIDIALYGTINEGVKGVGK 246

QY 246 TALKLIRHQSIESILENL--NKORYQIPEDWPYQEARRLFKEPNVTLDIPELKWTPADE 303  
DB 247 KALNYIKTGDIDFRALKALKKNIDH-----VEEIRNFELNPPVTDY-RIEFREPDF 297

QY 304 EGLISFLVKGDFNEDRVTKAEIKSAKNSSOGRLSEFF 344  
DB 298 EKAIEFLCEHDFSRERKEKALKAL--KSTQATLRF 336

RESULT 13  
H72765  
probable flap endonuclease-1 APE0115 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: H72765  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kawa, H.; 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: H72765  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-401 <KAW>  
A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79026.1; PID:g5103505  
A:Experimental source: strain K1  
C:Genetics:  
C:Gene: APE0115  
C:Superfamily: conserved hypothetical protein YKL113c

Query Match 26.5%; Score 514.5; DB 2; Length 401;  
Best Local Similarity 34.6%; Pred. No. 4.7e-27;  
Matches 119; Conservative 74; Mismatches 132; Indels 19; Gaps 6;

QY 15 PKAMKEQKESYFGKRIADVASMSIYQELIVVGRGTMETLTNEAGEVTSKLOGMENRTIR 74  
DB 62 PEARREVELRUGSVGLDALDAYNMLYQFLTAIRQPDGTPLDREGRVTSKLSGLFYRTIN 121

QY 75 LLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDATKDLTEAVEVGDKDAIEKLSKRTVKV 134  
DB 122 LVEEGIKPVYVFDGKPPDMKKQELAKRYSKRDATKDLTEAVEVGDKDAIEKLSKRTVKV 181

QY 135 TRHNECKRLRLMGVVPVVEAPSAEAEACALCINDKVFVASEDMDSITFGAPFLRLH 194  
DB 182 TSDMVEESKELLDAMGMPWQAPAGAEQAAYMARKDAGATWQYDSLLFGSPRLVKN 241

QY 195 LMDPSSKIPVME-----FDVAKVLELELTMDQFIDLCILGCDYCSIGIGG 244  
DB 242 LAITGRKLPGRDQYVEIKPEITIEPLLSKLTITREQLIANGILLGTIDYNGVGRVGYG 301

QY 245 QTALKLIRHQSIESILENLNKORYQIPEDWPY-QEARRLFKEPNVTLDIPELKWTPADE 303  
DB 302 KITALRLVSLGDMKVLASVPRGEY-----DPYLRLRVYVEFLNPPVTDY-KIEFRKPDQ 356

QY 304 EGLISFLVKGDFNEDRVTKAEIKSA---KNKSSOGRLSEFF 344  
DB 357 DKVREILVERHDFNPERVERALERLKGAYREKLRGQSLDMWF 400

RESULT 14  
F90158  
DNA repair endo/exonuclease FEN-1 (RAD2) (rad2) [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001;  
C:Accession: F90158  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-ngoc, H.P.; Redder, J.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: F90158  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-302 <KUR>  
A:Cross-references: GB:AE006641; NID:gl3813313; PIDN:AAK40525.1; GSPDB:GN00155  
C:Genetics:  
C:Gene: rad2  
C:Superfamily: conserved hypothetical protein YKL113c

Query Match 26.1%; Score 506; DB 2; Length 302;  
Best Local Similarity 37.9%; Pred. No. 1.2e-26;  
Matches 118; Conservative 58; Mismatches 97; Indels 38; Gaps 6;

QY 59 GVTSHLQGMNRTIRLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDATKDLTEAVEV 118  
DB 5 GVTSHLSGLFYRTINILEGVPIYVFDGKPEQKSEELERRRKAERKLERAKSE 64

QY 119 GDKDAIEKLSKRTVKVTRHNECKRLRLMGVVPVVEAPSAEAEACALCINDKVFVAVAS 178  
DB 65 GKIEELRYVSQALRLSNVVEESKLLRANGIPYVQAPSEGAEAAYLNKLGLSWAAS 124

QY 179 EDMDSLTGAPFLRLHMDPSKKIP-----VMEFDVAKVLELELTMDQFIDL 227  
DB 125 QYDAILFGAKRLVNLITITGKRLPNKVDYVEIKPELIEETI--LLKLGITR9QLIDI 182

QY 228 CILGCDY-CDSIKGIGGOTAKLIRHQSIESILE-----NLNKDRYQIPEDWPF 277  
DB 183 GILIGTDYNDGIRGIGPERALKIYKYEKAMEYGEISKDINFID----- 232

QY 278 QEARRLFKEPNVTLDIPELKWTPADEGLISFLVKGDFNEDRVTKAEIK-----IKSAKN 333  
DB 233 -BIRGLFNQVVPKPEEALDLNPNEDGIIINILVYEHNFSEERKNGIERLTKAIKEARG 291

QY 334 KSSQGRLESEFF 344  
DB 292 ASRQTCLDRWF 302

RESULT 15  
F84290  
DNA repair protein [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: F84290  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, J.; Leichner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jaung, K.H.; Alam, M.; Freitas, P.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.,  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: F84290







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:02:17 ; Search time 35 seconds  
(without alignments)  
449.129 Million cell updates/sec

Title: US-09-805-311-2

Perfect score: 1939

Sequence: 1 MGIKGLTKLLADNAPKAMKE.....SDKTSKAAANKTKAGGKKK 379

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1032	53.2	380	1	FENL_HUMAN
2	1023.5	52.8	380	1	RAD2_SCHPO
3	1014	52.3	378	1	FENL_MOUSE
4	938	48.4	382	1	RA27_YEAST
5	560.5	28.9	326	1	FEN_METJA
6	260.5	13.4	571	1	EX01_SCHPO
7	245.5	12.7	496	1	YA31_SCHPO
8	232.5	12.0	1031	1	RAD2_YEAST
9	220.5	11.4	1186	1	XPG_HUMAN
10	220	11.3	1170	1	XPG_MOUSE
11	218	11.2	702	1	EX01_YEAST
12	204.5	10.5	430	1	DIN7_YEAST
13	204.5	10.5	1112	1	RAL3_SCHPO
14	200	10.3	1196	1	XPG_XENLA
15	184.5	9.5	834	1	DP01_THECA
16	176.5	9.1	886	1	DP01_SYNT3
17	173	8.9	832	1	DP01_THERAQ
18	172.5	8.9	834	1	DP01_THERAQ
19	166	8.6	289	1	EX53_AQUAE
20	157.5	8.1	831	1	DP0F_THERH
21	156.5	8.1	833	1	DP01_THERH
22	154.5	8.0	877	1	DP01_STREP
23	153.5	7.9	877	1	DP01_LACILC
24	151.5	7.8	759	1	YENL_YEAST
25	149	7.7	930	1	DP01_HAEIN
26	144	7.4	877	1	DP01_BACCA
27	142.5	7.3	877	1	DP01_LACIA
28	138	7.1	921	1	DP01_RICHE
29	135.5	7.0	880	1	DP01_BACSU
30	134.5	6.9	856	1	DP01_DEIRA
31	134	6.9	867	1	DP01_RICPR
32	134	6.9	922	1	DP01_RICFE
33	123	6.3	876	1	DP01_BACST

RESULT 1  
FENL\_HUMAN  
AC P39748; 112892 seqs, 41476328 residues  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE FLAP endonuclease-1 (Maturation factor 1) (MFL).  
GN FENL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94277093; PubMed=8007985;  
RA Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S.,  
RA Lehmann A.R., Carr A.M., Watts F.Z.,  
RT "Structural and functional conservation of the human homolog of the  
RT Schizosaccharomyces pombe rad2 gene, which is required for chromosome  
RT segregation and recovery from DNA damage";  
RL Mol. Cell. Biol. 14:4878-4888(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX Tissue-Lymphocytes;  
RA Lieber M.R., Hsieh C.L., Harrington J.J.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX Lanerding J.E., McGready P.M., Coleman M., Skowronski E., Adamson A.W.,  
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,  
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,  
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,  
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,  
RA Cosfield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,  
RA Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,  
RA Kobayashi A., Olsen A.S.,  
RA Carraro A.V.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RX Tissue-Lung;  
RA Strausberg R.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP PARTIAL SEQUENCE.  
RX MEDLINE=95050647; PubMed=7961795;  
RA Robins P., Pappin D.J., Wood R.D., Lindahl T.;  
RT "Structural and functional homology between mammalian DNase IV and  
RT the 5'-nuclease domain of Escherichia coli DNA polymerase I";  
RL J. Biol. Chem. 269:28535-28538(1994).  
CC -!- FUNCTION: ENDONUCLEASE THAT CLEAVE 5'FLAP STRUCTURE AND FAILS  
CC TO CLEAVE OTHER DNA STRUCTURES, INCLUDING 3'FLAPS AND SINGLE  
CC STRANDED DNA (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. FENL

34 118.5 6.1 850 1 DP01\_ANATH  
35 113.5 5.9 347 1 ACPO\_CHLAW  
36 112 5.8 984 1 ACPO\_PEA  
37 112 5.8 984 1 SECA\_AQUAE  
38 109.5 5.6 891 1 DP01\_HELPY  
39 109.5 5.6 1016 1 DP01\_RHILE  
40 109.5 5.6 1650 1 BP28\_CABEL  
41 109.5 5.6 886 1 RA50\_ARCFU  
42 109 5.6 897 1 DP01\_HELPJ  
43 108.5 5.6 444 1 TIG\_CAMJE  
44 108.5 5.6 928 1 DP01\_ECOLI  
45 108.5 5.6 1509 1 MYSN\_ACACA

#### ALIGNMENTS

Q50156 anaerocellu  
O08307 chloroflexu  
P31239 pismu sativ  
C67718 aquifex aeo  
P56105 helicobacte  
Q98192 rhizobium l  
Q23495 caenorhabd  
O29230 archaeoglob  
Q921e9 helicobacte  
Q46108 campylobact  
P00582 escherichia  
P05659 acanthamoeb

CC CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X76771; CAAS4166.1; -

CC EMBL; L37374; AAA91331.1; -

CC EMBL; AC004770; AAC23394.1; -

CC EMBL; BC000323; AAH00323.1; -

CC HSSP; Q58839; 1A76.

CC Genew; HGNC:3650; FEN1.

CC MIM; 600393; -

CC InterPro; IPR002421; 5.3 exonuclease.

CC InterPro; IPR000513; Exo\_N.1.

CC InterPro; IPR003584; HHH\_2.

CC InterPro; IPR001532; XPGC\_Rad.

CC Pfam; PF00752; XPG\_N; 1.

CC Pfam; PF00867; XPG\_I; 1.

CC Pfam; PF00853; XPG\_RADSUPER.

CC SMART; SM00279; Hh2; 1.

CC SMART; SM00484; XPG1; 1.

CC SMART; SM00485; XPGN; 1.

CC PROSITE; PS00841; XPG\_1; 1.

CC PROSITE; PS00842; XPG\_2; 1.

CC Hydrolase; Nuclease; Endonuclease; Nuclear protein.

CC DOMAIN 1 104 N-DOMAIN.

CC FT DOMAIN 122 253 I-DOMAIN.

CC SEQUENCE 380 AA; 42593 MW; 5154F2F6E57592C5 CRC64;

CC -----

CC Query Match 53.28; Score 1032; DB 1; Length 380;

CC Best Local Similarity 53.88; Pred No. 2.2e-60;

CC Matches 206; Conservative 69; Mismatches 98; Indels 10; Gaps 6;

CC

CC QY 1 MGKGLTKLLADNAPKAMKEOKFESTFGKRIADVDSMSIYQPLIVVGTGMEITNAGE 60

CC DB 1 MGKGLTKLLADNAPKAMKEOKFESTFGKRIADVDSMSIYQPLIVVGTGMEITNAGE 60

CC QY 61 VTSHLQGFNRTIRLLEAGIKPVYVFDGKPPDMKKOELAKRYSKRDDATKDLTAEVGD 120

CC DB 60 TTSHLGMGFYRTIRMENGIKPVYVFDGKPPQLKSGELAKRSERRAEKOLQQAAGA 119

CC QY 121 KDAIEKLSKRTVYKTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFASVED 180

CC DB 120 EQEVEKFKRLVYKTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFASVED 179

CC QY 181 MSLLTFGAPRFLHMDPSKKIPVMEFDYAKVLELELTMDQFTDLGILGCDYCDISIK 240

CC DB 180 MCLTFGSPVLMRHLTASEAKPLQEFHLSRLQLGELNQEFVDFCLLGLSDYCESIR 239

CC QY 241 GIGGQTALKIRHGSTESILENKDRYQIPEDNPYQEARLFKEPNVTLID--IPELKW 298

CC DB 240 GIGPRKAVDLIOKHKSIEIIVRLDPNKPYPENWLNHKAHQLFLEPEV-LDPESVELKW 298

CC QY 299 TAPDEGLTSFLVKGNGNEDRVTKAJEIKSAKNSKSGRLESFKEPTATTSAPLRKE 358

CC DB 299 SEPNEELIKFVGEKQFSEERISRGVYKLSKRSQGSTQGRLLDFKVTGSLSS-AKRKE 357

CC QY 359 TSDK--TSKAAANKTKAGKKK 379

CC DB 358 PEKPGSTKKA---KTGAAGFK 377

CC

CC RESULT 2

CC RAD2\_SCHPO STANDARD; PRT; 380 AA.

CC ID RAD2\_SCHPO

CC AC P39750;

CC DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE DNA repair protein rad2.

GN RAD2 OR SPAC3G6.06C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94277093; PubMed=8007985;

RA Murray J.M., Tavassoli M., Al-Harithy R., Sheldrick K.S.,

RA Lehmann A.R., Carr A.M., Watts F.Z.;

RT "Structural and functional conservation of the human homolog of the

RT Schizosaccharomyces pombe rad2 gene, which is required for chromosome

RT segregation and recovery from DNA damage.";

RL Mol. Cell. Biol. 14:4878-4888(1994).

RN [2]

RP SEQUENCE FROM N.A.

RX STRAIN=972;

RA MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Soutos J., Peat N., Hayles J., Baker S., Basham D., Bowman C.M.,

RA Collins M., Connor R., Brown S., Chillingworth T., Churcher C.M.,

RA Gentles S., Goble A., Hamlin N., Harris D., Feltwell T., Fraser A.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA Jones K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,

RA Widdjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Leherer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Leherer H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Rhode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";

RL Nature 415:871-880(2002).

CC -1- FUNCTION: REQUIRED FOR THE FIDELITY OF CHROMOSOME SEPARATION AT

CC MITOSIS AND WHICH IS ALSO INVOLVED IN THE RESPONSE TO DNA DAMAGE.

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. FEN1

CC SUBFAMILY

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X77041; CAB36991.1; -

CC EMBL; Z99167; CAB16282.1; -

CC HSSP; Q58839; 1A76.

CC InterPro; IPR000513; Exo\_N.1.

CC InterPro; IPR003584; HHH\_2.

CC InterPro; IPR001532; XPGC\_Rad.

CC Pfam; PF00752; XPG\_N; 1.

CC Pfam; PF00867; XPG\_I; 1.

CC Pfam; PF00853; XPG\_RADSUPER.

CC SMART; SM00279; Hh2; 1.

CC SMART; SM00484; XPG1; 1.

DR SMART: SM00485; XPGN: 1.  
 DR PROSITE: PS00841; XPG\_1; 1.  
 DR PROSITE: PS00842; XPG\_2; 1.  
 KW DNA repair; Hydroxylase; Nuclease; Endonuclease; Nuclear protein.  
 FT DOMAIN 1 105 N-DOMAIN.  
 FT DOMAIN 123 254 I-DOMAIN.  
 SQ SEQUENCE 380 AA; 42866 MW; 3604065D9B934CBF CRC64;

Query Match 52.8%; Score 1023.5; DB 1; Length 380;  
 Best Local Similarity 50.7%; Pred. No. 8e-60;  
 Matches 194; Conservative 73; Mismatches 107; Indels 9; Gaps 3;

QY 1 MGIGLTKLLADNAPKAMEQKESYFGKRTAVDASMSIYQFLIVVGRGTMETLNEAGE 60  
 Db 1 MGIGLAQVLSEAPASVKHNDIKNIFGKRVDAIDASMSLYQFLIVRSODGQOLNEAGE 60  
 QY 61 VTSHLQGMENRTIRLEAGIKPVYVDFGKPPDMKKOELAKRYSKRDATKDLTEAVEVGD 120  
 Db 61 TTSHLGMENRTIRVNDGKPCFVDFGKPPDLKSGELAKRVARHOKAREDOEETKEVGT 120  
 QY 121 KDAIEKLSKRTVKTQHNEDCKRLRLMLGVVPEAPSEAEACALCINDKVFVASED 180  
 Db 121 AEMVDFRAKTKVKTQHNEDCKRLRLMLGVVPEAPSEAEACALCINDKVFVASED 180  
 QY 181 MDSLTGAPRFLRHLMDPSSKKIPVMEFDVAKLEELTMDQFIDLCLCGCDYCSIK 240  
 Db 181 MDLTGAPRFLRHLMDPSSKKIPVMEFDVAKLEELTMDQFIDLCLCGCDYCEPIR 240  
 QY 241 GIGGTALKLIRHGSISILENLKNDRYQIPEDWPYQEARLFPKPNVTLDP---EL 296  
 Db 241 GVGPARAVELIRGYTLDFRVKADRSKYPIDPWPYEDARRFLDAEV---LPGEETEL 297  
 QY 297 KWTAPDEGLISFLVKGDNFNDRTVKAIEKIKSAKNKSSQGRLESFFPKPTATTAPLKR 356  
 Db 298 KWSKSPDADGIIQFLVKEKGFNDRTVKAIEKIKSAKNKSSQGRLESFFPKPTATTAPLKR 356  
 QY 357 KETSDTKSKAAANKTKAGGKK 379  
 Db 356 VDTKSGSAKRKDSNKGESKK 378

RESULT 3  
 FEN1\_MOUSE  
 ID FEN1\_MOUSE STANDARD; PRT; 378 AA.  
 AC P39749;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE FLAP endonuclease-1.  
 GN FEN1 OR FEN-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 260-275 AND 331-352.  
 RC STRAIN=BALE/c;  
 RX MEDLINE=95011546; PubMed=7926735;  
 RA Harrington J.J., Lieber M.R.;  
 RT "Functional domains within FEN-1 and RAD2 define a family of  
 structure-specific endonucleases: Implications for nucleotide excision  
 repair.";  
 RL Genes Dev. 8:1344-1355(1994).  
 CC -!- FUNCTION: ENDONUCLEASE THAT CLEAVE 5' FLAP STRUCTURE AND FAILS  
 CC TO CLEAVE OTHER DNA STRUCTURES, INCLUDING 3' FLAPS AND SINGLE  
 CC STRANDED DNA.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. FEN1  
 CC SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L26320; AAC37664.1; -;  
 DR HSP; Q58839; IAY6.  
 DR MGD; MGI:102779; Fen1.  
 DR InterPro: IPR002421; 5\_3\_exonuclease.  
 DR InterPro: IPR000513; Exo\_N.1.  
 DR InterPro: IPR003584; HHH 2.  
 DR InterPro: IPR001532; XPGC\_Rad.  
 DR Pfam: PF00752; XPG\_N; 1.  
 DR Pfam: PF00867; XPG\_I; 1.  
 DR PRINTS; PR00853; XPGPADSUPER.  
 DR SMART; SM00279; HHH2; 1.  
 DR SMART; SM00484; XPG1; 1.  
 DR SMART; SM00485; XPGN; 1.  
 DR PROSITE; PS00841; XPG\_1; 1.  
 DR PROSITE; PS00842; XPG\_2; 1.  
 KW Hydroxylase; Nuclease; Endonuclease; Nuclear protein.  
 FT DOMAIN 1 102 N-DOMAIN.  
 FT DOMAIN 120 251 I-DOMAIN.  
 SQ SEQUENCE 378 AA; 42314 MW; 827946BD8BDEF39 CRC64;

Query Match 52.3%; Score 1014; DB 1; Length 378;  
 Best Local Similarity 53.2%; Pred. No. 3.3e-59;  
 Matches 202; Conservative 69; Mismatches 101; Indels 8; Gaps 6;

QY 1 MGIGLTKLLADNAPKAMEQKESYFGKRTAVDASMSIYQFLIVVGRGTMETLNEAGE 60  
 Db 1 MGIGLAQVLSEAPASVKHNDIKNIFGKRVDAIDASMSIYQFLIVRSODGQOLNEAGE 59  
 QY 61 VTSHLQGMENRTIRLEAGIKPVYVDFGKPPDMKKOELAKRYSKRDATKDLTEAVEVGD 120  
 Db 60 TTS-LMGNFYRTIR-MENGIRKPVYVDFGKPPDLKSGELAKRSERRAEKQLQQAEGM 117  
 QY 121 KDAIEKLSKRTVKTQHNEDCKRLRLMLGVVPEAPSEAEACALCINDKVFVASED 180  
 Db 118 EEEVEKTKRLVKTQHNEDCKRLRLMLGVVPEAPSEAEACALCINDKVFVASED 177  
 QY 181 MDSLTGAPRFLRHLMDPSSKKIPVMEFDVAKLEELTMDQFIDLCLCGCDYCSIK 240  
 Db 178 MDCLTFGSPVLMRHLTASEAKKLPIDQEPHLSRVLQELGQEQFVLDCLILGSDYCESIR 237  
 QY 241 GIGGTALKLIRHGSISILENLKNDRYQIPEDWPYQEARLFPKPNVTLDP---EL 299  
 Db 238 GIGAKRAVDLIQKHISIEIVRRLDPSPKYPVPEKNLHKEAQLLEPEVVDPESE/LKWS 297  
 QY 300 APDEGLISFLVKGDNFNDRTVKAIEKIKSAKNKSSQGRLESFFPKPTATTAPLKR 359  
 Db 298 EPNEEELVPMCKGEKQFSEIRIRSGVKRLSKSGROGTOGRDLDDFFKVTGSLSS-AKRKEP 356  
 QY 360 SDKTSKAAANKTKAGGKK 379  
 Db 357 E---PKGPAKKKAKTGAGK 373

RESULT 4  
 RAD2\_YEAST  
 ID RAD2\_YEAST STANDARD; PRT; 382 AA.  
 AC P26763;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Structure specific endonuclease RAD27.  
 GN RAD27 OR RTH1 OR YKL113C OR YKL510.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]

```
RP
RX MEDLINE-92221689; PubMed=1561835;
RA "Sequence of a 10.7 kb segment of yeast chromosome XI identifies the
RT APN1 and the RAF1 loci and reveals one tRNA gene and several new open
RT reading frames including homologs to RAD2 and kinases."
RL Yeast 8:121-132(1992).
RN [2]
RP CHARACTERIZATION.
RC STRAIN-BALB/C;
RX MEDLINE-95011546; PubMed=7926735;
RA Harrington J.J., Lieber M.R.;
RT "Functional domains within FEN-1 and RAD2 define a family of
RT structure-specific endonucleases: implications for nucleotide
RT excision repair."
RL Genes Dev. 8:1344-1355(1994).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE-95113773; PubMed=7814325;
RA Reagan M.S., Pittenger C., Siede W., Friedberg E.C.;
RT "Characterization of a mutant strain of Saccharomyces cerevisiae with
RT a deletion of the RAD27 gene, a structural homolog of the RAD2
RT nucleotide excision repair gene."
RL J. Bacteriol. 177:364-371(1995).
CC -1- FUNCTION: SINGLE-STRANDED DNA ENDONUCLEASE (PROBABLE).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. FEN1
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S93804; AAB21998.1;
DR EMBL: Z28113; CAAB1953.1;
DR PIR: S22267; S22267.
DR HSP: Q58839; 1A76.
DR SGD: S0001596; RAD27.
DR InterPro: IPR002421; 5_3_exonuclease.
DR InterPro: IPR005013; EXO_N1.
DR InterPro: IPR003384; HHH_2.
DR InterPro: IPR001532; XPGC_Rad.
DR Pfam: PF00752; XPG_N; 1.
DR Pfam: PF00867; XPG_I; 1.
DR Pfam: PF01367; 5_3_exonuclease; 1.
DR PRINTS: PR00853; XPG_RADSUPER.
DR SMART: SM00279; HHH2; 1.
DR SMART: SM00484; XPG1; 1.
DR SMART: SM00485; XPGN; 1.
DR PROSITE: PS00841; XPG_1; 1.
DR PROSITE: PS00842; XPG_2; 1.
KW DNA repair; Hydrolase; Nuclease; Endonuclease; Nuclear protein.
FT DOMAIN 1 105 N-DOMAIN.
FT DOMAIN 120 251 I-DOMAIN.
SQ SEQUENCE 382 AA; 43279 MW; 1F54B08720121C8C CRC64;

Query Match 48.4%; Score 938; DB 1; Length 382;
Best Local Similarity 49.5%; Pred. No. 3e-54;
Matches 190; Conservative 69; Mismatches 109; Indels 16; Gaps 6;

Qy 1 MGKGLTKLIADNAPKAKKEKFESYFGKRIAYDASMSYQFLIVVGRGTGMEITLNEAGE 60
Db 1 MGKGLNATISEHVPISATKSDIKSFGRKVAIDASMSYQFLIVVGRGTGMEITLNEAGE 60
Qy 61 VTSHLGGMNRTIRLEAGIKPVVYFDGKPPDMKKOELAKRYSKRDKATKDLTAVEVGD 120
Db 61 TTSHLGGMNRYRLMDINGIRFCYFDGKPPDLKSHLTKRSRRVVEKKLAETTELE 120
Qy 121 KDAIEKLSKRTVKVTRQNEDECKRLRLMGVPVVEAPSEAECAALCINDKVFVASED 180

SEQUENCE FROM N.A.
121 K--MKOERRLVKSKHEHNEAQLGLMGIPYTIAPTEAEQAQCAELAKKGVVAAASD 177
Qy 181 MDSITFCAPFRLHLMDSPPSKKIPVMEFVAKVLELELTMDQFDLCTLCGLGVCDSIK 240
Db 178 MDVLCTPFLRLHRTFSKAKKEPIHEIDTFLVGLDITIEQFVDCIMLGCDCYCESIR 237
Qy 241 GIGGQTKALKIRHOGHSIESILENL-----NKDRYQIPEDPWPQEARLKEPNVTLDIPE 295
Db 238 GVGPTVTKALKIRHOGHSIESIKVIEFIESGESNNTRKWKIPEDWPYKQARMFLDPEV-IDGNE 296
Qy 296 --LKVTPADEGLISFLYKDNFGNEDRVTKAIEKIKSAKNKSQGRLESFFKPTATT--- 350
Db 297 INLKWSPPEKKEIIEYLCDDKFKSEERVKSGISRLAKGLKGLTQGRDLGDFQFVVPKTEQ 356
Qy 351 -SAPLRKRKETSDKTSKAAANKTK 373
Db 357 LAAAKRAQENKKLNK-NKNKVTK 379

RESULT 5
FEN_MENJA STANDARD; PRT; 326 AA.
ID FEN_MENJA
AC Q58839;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Flap endonuclease.
GN FEN OR MJ1444.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE-96337999; PubMed=6688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.B., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii."
RT Science 273:1058-1073(1996).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE-98440432; PubMed=9765572;
RA Rao H.G., Rosenfeld A., Wetmur J.G.;
RA "Methanococcus jannaschii flap endonuclease: expression, purification,
RT and substrate requirements."
RL J. Bacteriol. 180:5406-5412(1998).
RN [3]
RP CHARACTERIZATION, AND X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE-99200693; PubMed=10102570;
RA Bae K.W., Bae K.W., Cho C.S., Hwang K.Y., Kim H.-R., Sung H.-C.,
RA Cho Y.;
RA "Expression, purification, characterization and crystallization of
RT flap endonuclease-1 from Methanococcus jannaschii."
RL Mol. Cells 9:45-48(1999).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE-98363215; PubMed=9699635;
RA Hwang K.Y., Bae K., Kim H.-Y., Cho Y.;
RA "The crystal structure of flap endonuclease-1 from Methanococcus
RT jannaschii."
RL Nat. Struct. Biol. 5:707-713(1998).
CC -1- FUNCTION: Has 5' endo-/exonuclease and 5' pseudo-Y-endonuclease
CC activities. Cleaves the junction between single and double-
CC stranded regions of flap DNA.
```

```

CC CC      -!- COFACTOR: Binds 2 magnesium ions per subunit.
CC CC      -!- MISCELLANEOUS: ITS OPTIMUM PH IS 6-7.
CC CC      -!- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. FEN1
CC CC      SUBFAMILY.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL: U67585; AAB99454.1;
CC CC      PDB: 1A76; 03-AUG-99.
CC CC      PDB: 1A77; 03-AUG-99.
CC CC      TIGR: MJ1444.
CC CC      InterPro: IPR000513; Exo_N_I.
CC CC      InterPro: IPR001532; XPGC_Rad.
CC CC      Pfam: PF00752; XPG_N; 1.
CC CC      Pfam: PF00867; XPG_I; 1.
CC CC      PRINTS: PR00853; XPGRADSUPER.
CC CC      SMART: SM00279; HH2; 1.
CC CC      SMART: SM00484; XPGI; 1.
CC CC      SMART: SM00485; XPGN; 1.
CC CC      Hydroxylase; Nuclease; Endonuclease; Magnesium; Metal-binding;
CC CC      3D-structure; Complete proteome.
CC CC      METAL 154 MAGNESIUM 1.
CC CC      SEQUENCE 326 AA; 37527 MW; 311427F2B4B67580 CRC64;
CC CC      -----
CC CC      Query Match 28.9%; Score 560.5; DB 1; Length 326;
CC CC      Best Local Similarity 39.1%; Pred. No. 9.5e-30;
CC CC      Matches 127; Conservative 68; Mismatches 114; Indels 16; Gaps 7;
CC CC      -----
CC QY 23 FESYFGKIAVDASMSIYQFLIVGRTGMETLTNEAGEVTSHQGMFETILLIAGTKP 82
CC Db 16 FEDLKKKKVAIDGMNLYQFLSIRLDSGSRNKRKGEIYSANVGYFTIHLLENDITP 75
CC QY 83 VYVFDGKPPDKQKQELAKRYKSRDDATKDLTEAVEVDKDAIKELSKRVKVTROHNEDC 142
CC Db 76 IIVFDGEPKPKKTRKVRKREKAKELKREAKKEDEFAAKYAKRVSYLTPKRVENC 135
CC QY 143 KELLRLMGVPVPAEAEACALCINDKVFVAVASEMDLSLFTGAPRFLRLMDPSSK 202
CC Db 136 KYLLSLMGIPVPAEAGEAQASVMAKKGVDYVAVSQDYDALLYGAPRVYRNL--TTTKE 193
CC QY 203 IPVMEFDVAKVLELTMDQFIDLCILCGDY-CDSIKIGGQTKALKIRHQHSIESIL 261
CC Db 194 MPFL-IELNEVLDELRLISLDDLDLIAIFMGTDYNPNGVKGIGPKRAYELVRS-GVAKDVL 251
CC QY 262 ENLNKRDYQIPEDWPYQEARLFEPNVTLDIPELKWTPADEGLISFLVKONGFENEDRV 321
CC Db 252 KK-----EVEYDEIKRIFKPKVT-DNYSLSLKLDPKKEGIILKFLVDENDENYDRV 301
CC QY 322 TKALEIKTS-AKNKSQGRLESFFK 345
CC Db 302 KKHVDKLYLNANKTKQTLDAWFK 326
CC CC      -----
CC CC      RESULT 6
CC CC      EXOL_SCHPO
CC ID EXOL_SCHPO STANDARD: PRT; 571 AA.
CC AC P53695;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Exonuclease I (EXO I).
CC GN EXO1 OR SPBC29A10.05.
CC OS Schizosaccharomyces pombe (Fission yeast).
CC OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC OC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC OC Schizosaccharomycetes.

```

```

OX CC      NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 139-160.
RC STRAIN=972;
RX MEDLINE=95158876; PubMed=7855597;
RT "A role for exonuclease I from S. pombe in mutation avoidance and
RT mismatch correction."
RL Science 267:1166-1169(1995).
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat M., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeil C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharf S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Carzon A., Thode G.,
RA Dega R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe J., McCombie W.R., Paulsen I., Potashkin J.,
RA Spakovits G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC CC      -!- FUNCTION: 5'->3' DOUBLE-STRANDED DNA EXONUCLEASE THAT COULD ACT IN
CC CC      -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC CC      -!- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL: L35174; AAC1648.1;
CC CC      EMBL: AL034463; CA22433.1;
CC CC      InterPro: IPR000513; EXO_N_I.
CC CC      InterPro: IPR003584; HH2.
CC CC      InterPro: IPR001440; TPR.
CC CC      InterPro: IPR001532; XPGC_Rad.
CC CC      Pfam: PF00752; XPG_N; 1.
CC CC      Pfam: PF00867; XPG_I; 1.
CC CC      PRINTS: PR00853; XPGRADSUPER.
CC CC      SMART: SM00279; HH2; 1.
CC CC      SMART: SM00484; XPGI; 1.
CC CC      SMART: SM00485; XPGN; 1.
CC CC      PROSITE: PS00841; XPG_1; 1.
CC CC      PROSITE: PS00842; XPG_2; 1.
CC CC      DNA repair; Hydrolase; Nuclease; Endonuclease; Nuclear protein.
CC CC      DOMAIN 1 96
CC CC      DOMAIN 114 245 I-DOMAIN.
CC CC      SEQUENCE 571 AA; 63867 MW; 7D21B3778104BA9D CRC64;
CC CC      Query Match 13.4%; Score 260.5; DB 1; Length 571;

```

```

Best Local Similarity 25.4%; Pred. No. 6.6e-10;
Matches 110; Conservative 73; Mismatches 167; Indels 83; Gaps 17;

QY 1 MGKIGLTKLLADNAPAM-KEOKFESYFGRIKIAVDASMSIYQFLVVGRTGMEFLTNEAG 59
DB 1 MGKIGLLGLL-----KPMQKSHVVEESGKTLGVDGVWLHKAIVTCAH-----ELAFNK-- 50
QY 60 EVTSHLOGMEITIRLLIAGIKPVYVDSKPPDMKKQELAKRYSKRDATKDLTEAVEVG 119
DB 51 ETKYLKYLHQAALMQLYGVPLVDFDGGPLCKASTQKRRKQ-----EAFELG 103
QY 120 DK-----DAIEKLSKRTVKVTRQHNEDCKRLRLMVGVPVVEAPSAAEACALCIND 171
DB 104 KKLWDEKKSQAIMQFS-RCVDVTPEMAWKLIIALREHIGESIVAPYEAQAQLVLEKEN 162
QY 172 KVFASVEDMSLITGAPRFLRHLMDFSSKKIPVMEFDVAKYLE-ELBTLMDQFDILCIL 230
DB 163 IIDGITEDSDMLVFGAQTVL-FKMDGFGNCITIRNDIANAOQDLNLRPIEKLRHMAIF 221
QY 231 CGGDCYCDSTKIGGOTALKLIRHQGSTIESILENLKDR-YQIPEDWYOEAE-----R 281
DB 222 SGCDYTDGVAGMGLKTLRYLQKYPEPRAAIRAMRLDKSLKVPVSFEKEFALADLAFRHQ 281
QY 282 RLKPEPNVT---LDIPELKWTPADDEGLISLVKONGF-----NEDRVTKATEIKTSA 331
DB 282 RVYCPKDKTLVHLSPEPRLSVHEDAFIGSFF-DNQLAIDIAEGRSPNPIKCAFDIKDS 339
QY 332 KNKS-----SQGRLESFF-----KPTATTS-----APLKRKE 358
DB 340 SMOSFTKTTITIKRKGISKYTDISNFFMKSIPTPSKRPKTSLIDVTWVKVQRTHLANDI 399
QY 359 TSDKTSKAAANK 371
DB 400 SSKQSIKSANEK 412

RESULT 7
ID YA31_SCHPO STANDARD; PRT; 496 AA.
AC Q05708:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C18B11.01c in chromosome I.
GN SPAC18B11.01c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris B., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Coiffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

```

```

RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. FEN1
CC SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z50728; CAA90586.1;
CC InterPro; IPR000513; Exo_N.I.
CC InterPro; IPR003584; HHH_2.
CC InterPro; IPR001532; XPGC_Rad.
CC Pfam; PF00752; XPG_N; 1.
CC Pfam; PF00867; XPG_I; 1.
CC PRINTS; PR00853; XPGADSUPER.
CC SMART; SM00484; XPG1; 1.
CC SMART; SM00485; XPGN; 1.
CC SMART; SM00485; XPGN; 1.
CC Hypothetical protein. Hydrolase; Nuclease; Endonuclease.
CC KW SEQUENCE 496 AA; 57763 MW; FFA16D26846F24C3 CRC64;
CC -----
Query Match 12.7%; Score 245.5; DB 1; Length 496;
Best Local Similarity 30.2%; Pred. No. 5.3e-09;
Matches 62; Conservative 47; Mismatches 77; Indels 19; Gaps 5;

QY 111 DLTEAVEVGDKDAIEKLSKRTVKVTRQHNEDCKRLRLMVGVPVVEAP--SAAEACALC 168
DB 251 ELQAKIAKLTQTE-LDKLERLYRSPQNIFFELIKILGPASFPIGVEAEAFASIS 309
QY 169 INDKVFAVASEDMDSLTGAPRFLRHLMDFSSKKIPVMEFDVAKYLEELTMDQFDILC 228
DB 310 QNNLAYAVATQDVTLLGSSMISNFDLNDNFHLPLQIMDKIAQELNLTDFGFDYQC 369
QY 229 ILCCGDCYCDSTKIGGOTALKLIRHQGSTIESILENLN-KORYOIPEDM--PYOEARRLFK 285
DB 370 LMCQGFDTFSRIPKIGPVRLKIRYGNADFVRLKALNVEEKYIIPDYIKKFLTKKLFT 429
QY 286 EPNVTLDIPELKWTPADDEGLISFL 310
DB 430 D-----LPSNNELFSFI 441

RESULT 8
RAD2_YEAST
ID RAD2_YEAST STANDARD; PRT; 1031 AA.
AC P07276;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA repair protein RAD2.
GN RAD2 OR YGR258C.
OS Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=86083164; PubMed=3000874;
RA Nicolet C.M., Chenevert J.M., Friedberg E.C.;
RT "The RAD2 gene of Saccharomyces cerevisiae: nucleotide sequence and
RT transcript mapping.";
RL Gene 36:225-234(1985).
RN [2]

```



RA Clarkon S.G.;  
 RT "The human gene for Xeroderma pigmentosum complementation group G  
 RL (XPG) maps to 13q33 by fluorescence in situ hybridization.";  
 RN Genomics 21:283-285(1994).  
 [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=94266772; PubMed=8206890;  
 RA O'Donovan A., Scherly D., Clarkson S.G., Wood R.D.;  
 RT "Isolation of active recombinant XPG protein, a human DNA repair  
 RL endonuclease.";  
 RN J. Biol. Chem. 269:15965-15968(1994).  
 [6]  
 RP CHARACTERIZATION.  
 RX MEDLINE=94376899; PubMed=8090225;  
 RA O'Donovan A., Davies A.A., Moggs J.G., West S.C., Wood R.D.;  
 RT "XPG endonuclease makes the 3' incision in human DNA nucleotide  
 RL excision repair.";  
 RN Nature 371:432-435(1994).  
 [7]  
 RP CHARACTERIZATION.  
 RX MEDLINE=94359802; PubMed=8078765;  
 RA Habraken V., Sung P., Prakash L., Prakash S.;  
 RT "Human Xeroderma pigmentosum group G gene encodes a DNA  
 RL endonuclease.";  
 RN Nucleic Acids Res. 22:3312-3316(1994).  
 [8]  
 RP REVIEW ON VARIANTS XP-G.  
 RX MEDLINE=93374920; PubMed=10447254;  
 RA Cleaver J.E., Thompson L.H., Richardson A.S., States J.C.;  
 RT "A summary of mutations in the UV-sensitive disorders: Xeroderma  
 RL pigmentosum, Cockayne syndrome, and trichothiodystrophy.";  
 RN Hum. Mutat. 14:9-22(1999).  
 [9]  
 RP VARIANT XP-G VAL-792, AND VARIANT ASP-1104.  
 RX MEDLINE=95038755; PubMed=7951246;  
 RA Nussipikel T., Clarkson S.G.;  
 RT "Mutations that disable the DNA repair gene XPG in a Xeroderma  
 RL pigmentosum group G patient.";  
 RN Hum. Mol. Genet. 3:963-967(1994).  
 [10]  
 RP VARIANT XP-G VAL-792.  
 RX MEDLINE=97250499; PubMed=9096355;  
 RA Nussipikel T., Lalle P., Leadon S.A., Cooper P.K., Clarkson S.G.;  
 RT "A common mutational pattern in Cockayne syndrome patients from  
 RL Xeroderma pigmentosum group G: implications for a second XPG  
 function.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 94:3116-3121(1997).  
 CC -1- FUNCTION: SINGLE-STRANDED DNA ENDONUCLEASE INVOLVED IN DNA  
 CC EXCISION REPAIR. MAKES THE 3' INCISION IN DNA NUCLEOTIDE EXCISION  
 CC REPAIR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- DISEASE: COMPLEMENTS THE DEFECT IN XERODERMA PIGMENTOSUM (XP)  
 CC GROUP G (XP-G). AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY  
 CC HYPERSENSITIVITY OF THE SKIN TO SUNLIGHT FOLLOWED BY HIGH  
 CC INCIDENCE OF SKIN CANCER AND FREQUENT NEUROLOGIC ABNORMALITIES.  
 CC -1- DISEASE: ALSO INVOLVED IN COCKAYNE'S SYNDROME (CS); AN AUTOSOMAL  
 CC RECESSIVE DISEASE WHICH IS CHARACTERIZED BY A UV-SENSITIVE SKIN  
 CC (WITHOUT PIGMENTATION ABNORMALITIES), NEUROLOGICAL DYSFUNCTION  
 CC DUE TO DEMYELINATION OF NEURONS AND CALCIFICATION OF BASAL GANGLIA  
 CC (PSYCHOMOTOR RETARDATION, DEAFNESS, OPTIC ATROPHY, RETINAL  
 CC PIGMENTATION AND HYPERREFLEXES), AND DYSMORPHIC DWAFTISM (IMMATURE  
 CC SEXUAL DEVELOPMENT AND MICROCEPHALY).  
 CC -1- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. XPG  
 CC SUBFAMILY.  
 CC -1- DATABASE: NAME=Atlas Genet. Cytogetnet. Oncol. Haematol.;  
 CC WWW="http://www.infobiogen.fr/services/Chromancer/Genes/XPGID300.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC  
 CC  
 DR EMBL; X69978; CAA49598.1; -  
 DR EMBL; D16305; BAA03812.1; -  
 DR EMBL; L20046; AAC37533.1; -  
 DR EMBL; X71341; CAA50481.1; -  
 DR EMBL; X71342; CAA50481.1; JOINED.  
 DR FIR; S35993; S35993.  
 DR Genew; HGNC:3437; ERCC5.  
 DR MIM; 133530; -  
 DR MIM; 278780; -  
 DR InterPro; IPR000513; Exo\_N.1.  
 DR InterPro; IPR003584; HHH\_2.  
 DR InterPro; IPR001044; XPGC\_DNA\_repair.  
 DR InterPro; IPR001532; XPGC\_Rad.  
 DR Pfam; PF00752; XPG\_N; 1.  
 DR Pfam; PF00867; XPG\_I; 1.  
 DR PRINTS; PR00853; XPGRADSUPER.  
 DR SMART; SM00279; Hhh2; 1.  
 DR SMART; SM00484; XPGI; 1.  
 DR SMART; SM00485; XPGN; 1.  
 DR TIGRfam; TIGR00600; rad2; 1.  
 DR PROSITE; PS00841; XPG\_1; 1.  
 DR PROSITE; PS00842; XPG\_2; 1.  
 DR DNA repair; DNA-binding; Nuclear protein; Hydrolase; Nuclease;  
 KW Endonuclease; Xeroderma pigmentosum; Cockayne's syndrome;  
 KW Polymorphism; Disease mutation.  
 FT DOMAIN 1 95 N-DOMAIN.  
 FT DOMAIN 753 881 I-DOMAIN.  
 FT DOMAIN 1057 1073 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT VARIANT 254 254 M -> V.  
 FT VARIANT 792 792 /FTID-VAR\_007732.  
 FT VARIANT 1104 1104 A -> V (IN XP-G; MILD FORM).  
 FT VARIANT 1104 1104 H -> D.  
 FT CONFLICT 55 55 /FTID-VAR\_007734.  
 FT CONFLICT 120 122 L -> P (IN REF. 2).  
 FT CONFLICT 126 126 KTA -> QTS (IN REF. 2).  
 FT CONFLICT 126 126 K -> Q (IN REF. 2).  
 FT CONFLICT 264 266 ROY -> SSH (IN REF. 2).  
 FT CONFLICT 760 760 I -> F (IN REF. 2).  
 FT CONFLICT 796 796 I -> V (IN REF. 2).  
 FT CONFLICT 864 872 EGIPTGVCV -> GNTNGLC (IN REF. 2).  
 FT CONFLICT 959 959 R -> S (IN REF. 2).  
 SQ SEQUENCE 1186 AA; 133328 MW; 8702EC68E080F1C4 CRC64;  
 Query Match 11.4%; Score 220.5; DB 1; Length 1186;  
 Best Local Similarity 21.3%; Pred. No. 6.6e-07;  
 Matches 94; Conservative 71; Mismatches 149; Indels 127; Gaps 15;  
 QY 15 PKAMKQKPFESYGRKIAVDASMSIYQFLIVGRTGMEVLTNEAGEVTSHLQGMFNTIR 74  
 DB 635 PKAVEPWEIDS-----ESESDESGSFIEV-----QSVISDE-----ELQAEFPET-- 673  
 QY 75 LLEAGIKPVYVFDGKPPDMKQK-----LAKRYSKRDDATKDLTEAVEV 118  
 DB 674 -----SKPPEQEEELVTRGEAPAESESLLDNSERDDVDGEPQEA--- 717  
 QY 119 GDDDAIEKL-----SKRVKVTROHNE-----DCKRL 146  
 DB 718 -EKDAEDSLHEWODINLELETLESNLAQNSLAKQKQKQRIAAVVTGMFLESQELL 776  
 QY 147 RLGMGVFVPEAPSEAEACALCINDKVFVAVASEDMDSLTFFGAPRFLRLHMDPSSKIPVM 206  
 DB 777 RLFGIPYIQAPAEAEACAILDLDTSGTITDSDILWLFGARHYVRNFF---NKNKFE 833  
 QY 207 EFDVAVKVELELTMDQFDILCIGDCYCDISIKGIGTALKLIQ-----HGSTEILE- 262  
 DB 834 YYQYVDFHQLGLDRNKLINLAYLIGSDYEGIPVGCVTAMEILNEFPFGH-LEPLKX 892  
 QY 263 -----NLNKD--RYQIPEDWPYQEARLKFEPNVTLDIPELKWTA 300  
 DB 893 SEWHEAQNKPIRPNPHDHTVKKKKLTLOLTGFPNPAVAEAYLKPVVDSSKGSFLNGK 952



QY 301 PDEGLISPLVDKNGFNEDRVTKAI-EKISAKNKSQORLESFFKPTATTISAPLKRKET 359  
 Db 953 PDLDKIREPCQRYFWNTKTESLFPVLQADQOQLRIDSFFFLAQOKEKEDAKRIKS 1012  
 QY 360 S-----DKTSKAAANK 370  
 Db 1013 QRLNRAVTCMLRKEKAAASE 1033

RESULT 10  
 XPG\_MOUSE STANDARD; PRT; 1170 AA.  
 AC P35689; Q61528; Q64248;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA-repair protein complementing XP-G cells homolog (Xeroderma  
 DE pigmentosum group G complementing protein homolog) (DNA excision  
 DE repair protein ERCC-5).  
 GN ERCC5 OR XPG OR ERCC-5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP MEDLINE-94173288; PubMed-7510366;  
 RX Shiomi T., Harada Y.-N., Saito T., Shiomi N., Okuno Y., Yamaizumi M.;  
 RA "An ERCC5 gene with homology to yeast RAD2 is involved in group G  
 RT Xeroderma pigmentosum.";  
 RL Mutat. Res. 314:167-175(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/10; TISSUE=Liver;  
 RX MEDLINE-96070433; PubMed-7590748;  
 RA Harada Y.N., Matsuda Y., Shiomi N., Shiomi T.;  
 RT "Complementary DNA sequence and chromosomal localization of xpg, the  
 RL mouse counterpart of human repair gene XPG/ERCC5.";  
 RN Genomics 28:59-65(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DRA/2;  
 RX MEDLINE-96359149; PubMed-8703115;  
 RA Ludwig D.L., Mudgett J.S., Park M.S., Perez-Castro A.V.,  
 RA MacInnes M.A.;  
 RT "Molecular cloning and structural analysis of the functional mouse  
 RL genomic XPG gene.";  
 RN Mamm. Genome 7:644-649(1996).  
 CC -1- FUNCTION: SINGLE-STRANDED DNA ENDONUCLEASE INVOLVED IN DNA  
 CC EXCISION REPAIR. MAKES THE 3' INCISION IN DNA NUCLEOTIDE EXCISION  
 CC REPAIR (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. XPG  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D16306; BAA03813.1; -  
 CC EMBL: U40796; AAA91039.1; -  
 CC EMBL: U40795; AAB17885.1; -  
 CC EMBL: U39892; AAB17885.1; JOINED.  
 CC EMBL: U39893; AAB17885.1; JOINED.  
 CC EMBL: U39894; AAB17885.1; JOINED.  
 CC EMBL: U39895; AAB17885.1; JOINED.  
 CC EMBL: U40073; AAB17885.1; JOINED.  
 CC EMBL: U40431; AAB17885.1; JOINED.

DR EMBL: U40432; AAB17885.1; JOINED.  
 DR EMBL: U40668; AAB17885.1; JOINED.  
 DR EMBL: U40669; AAB17885.1; JOINED.  
 DR EMBL: U40670; AAB17885.1; JOINED.  
 DR EMBL: U40792; AAB17885.1; JOINED.  
 DR EMBL: U40793; AAB17885.1; JOINED.  
 DR EMBL: U40794; AAB17885.1; JOINED.  
 DR MGD; MGI:103582; Ercc5.  
 DR InterPro; IPR000513; EXO\_N\_I.  
 DR InterPro; IPR003584; HHH\_2.  
 DR InterPro; IPR001044; XPGC\_DNA\_repair.  
 DR InterPro; IPR001532; XPGC\_Rad.  
 DR Pfam; PF00752; XPG\_N; 1.  
 DR PRINTS; PR00867; XPG\_I; 1.  
 DR SMART; SM00279; HhH2; 1.  
 DR SMART; SM00484; XPGI; 1.  
 DR SMART; SM00485; XPGN; 1.  
 DR TIGRFAMs; TIGR00600; rad2; 1.  
 DR PROSITE; PS00841; XPG\_1; 1.  
 DR PROSITE; PS00842; XPG\_2; 1.  
 DR DNA repair; DNA-binding; Nuclear protein; Hydrolase; Nuclease;  
 KW Endonuclease.  
 FT DOMAIN 1 95 N-DOMAIN.  
 FT DOMAIN 752 882 I-DOMAIN.  
 FT DOMAIN 1049 1065 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT VARIANT 388 488 R -> C (IN STRAIN DBA/2).  
 FT VARIANT 488 488 S -> R (IN STRAIN DBA/2).  
 FT VARIANT 688 688 I -> T (IN STRAIN DBA/2).  
 FT VARIANT 1015 1015 S -> T (IN STRAIN DBA/2).  
 FT VARIANT 1021 1021 M -> I (IN STRAIN DBA/2).  
 FT VARIANT 1121 1121 S -> P (IN STRAIN DBA/2).  
 FT CONFLICT 227 227 N -> M (IN REF. 1).  
 FT CONFLICT 249 249 N -> M (IN REF. 1).  
 FT CONFLICT 300 302 VMD -> MDE (IN REF. 1).  
 FT CONFLICT 313 313 N -> M (IN REF. 1).  
 FT CONFLICT 320 320 N -> M (IN REF. 1).  
 FT CONFLICT 399 399 D -> DVDTGG (IN REF. 1).  
 FT CONFLICT 408 408 N -> NSASEVIGPV (IN REF. 1).  
 FT CONFLICT 581 581 M -> V (IN REF. 1).  
 FT CONFLICT 795 795 N -> A (IN REF. 1).  
 FT CONFLICT 1039 1039 A -> AMEKEFEL (IN REF. 1).  
 FT CONFLICT 1134 1134 S -> SD (IN REF. 1).  
 FT CONFLICT 1157 1158 KL -> RR (IN REF. 1).  
 SQ SEQUENCE 1170 AA; 130864 MW; 4058F07FFD1770ED CRC64;  
 Query Match 11.3%; Score 220; DB 1; Length 1170;  
 Best Local Similarity 21.6%; Pred. No. 7e-07;  
 Matches 88; Conservative 71; Mismatches 156; Indels 92; Gaps 13;  
 QY 13 NAPKAMKEKTESYFGRIKIAVDASMSIYQFLIYVGRGTMETLTNEAGVTSHL----- 65  
 Db 633 SAKPMGPMEMES-----ESESOGSFIQSVSVNSSELOTESSEASHLSEKDAEE 684  
 QY 66 -----QGMFNRTLLRLEAGIKPYVVDGKPPDMKKQELAKRYSKRDDATKDLT---EAV 116  
 Db 685 PREILEEGYSRDETECLQ-----DSSDIEAMEGHEGREADDAEDMNEWQDI 730  
 QY 117 EYGDKDAIEK---LSKRTVKVKTROHNE-----DCKRLRLMGVZVVPAPSEAE 161  
 Db 731 NLEELDALESNLLARQNSLKAQKQODRIAASVTGOMFLESQELLRLFGVPTICAPMEAE 790  
 QY 162 AECAALCINDKVFVAVASDMDSITFGAPRFLRHLMDPSSKIPVME-FDVAKVLEELELT 220  
 Db 791 AQACMLDLSQTSGITDSDINLFGA---RHVYKKNFNKNKKEVYQYVDFYSQLGLD 846  
 QY 221 MOOFTDLCILGCDYCSIKIGGOTALKIRQH--GSIESIL-----ENLKN--- 266  
 Db 847 RNKLINLAYLGSYTEGIPYGVCTAMEINFEPPGRLDPLLKFESEKHWQAQNKVAAE 906  
 QY 267 -----DRYQIPEDWPYQEARLFPKPNVTLDIPELKWTAPDEEGLISFVXDNG 315  
 Db 907 NPYDTKVRKKLRKLQLTQPGFPNPAVADAYLRPVYDDSRGSLFKGPKPDYDKIRFXQRYFG 966

QY 316 FNEEDRVTRAIKBI-KSAKNSSQGRLESFPKPTATTSAPLKRKETS 361  
 Db 967 WNRMTDESPLVPLKHLNAHOTLRIDSFR-----LAQQRKQD 1005

RESULT 11  
 ID EXO\_YEAST STANDARD; PRT; 702 AA.  
 AC P39875;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Exonuclease I (EXO I) (DHS1 protein).  
 GN EXO1 OR DHS1 OR YOR033C OR OR26.23.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE OF 220-702 FROM N.A.  
 RC STRAIN=S288c / FY1679;  
 RA de Haan M., Maarse A.C., Grivell L.A.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Landt O., Hiesel R., Unseld M.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Tishkoff D.X., Boeger A.L., Filosi N., Galda G.M., Bertrand P.,  
 RA Kane M.F., Kolodner R.D.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP FUNCTION  
 RX MEDLINE=97265410; PubMed=9111347;  
 RA Fiorentini P., Huang K.N., Tishkoff D.X., Kolodner R.D.,  
 RA Symington L.S.;  
 RA "Exonuclease I of Saccharomyces cerevisiae functions in mitotic  
 recombination in vivo and in vitro.";  
 RL Mol. Cell. Biol. 17:2764-2773(1997).  
 CC -!- FUNCTION: 5'-3' double-stranded DNA exonuclease involved in  
 mismatch repair and eventually also in mitotic recombination  
 between direct repeats.  
 CC -!- COFACTOR: Magnesium or manganese.  
 CC -!- ENZYME REGULATION: Inactivated by calcium and zinc ions.  
 CC -!- SUBUNIT: INTERACTS WITH MISMATCH REPAIR PROTEIN MSH2.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; S69545; AAC60570.1;  
 CC EMBL; X87331; CAA60749.1;  
 CC EMBL; 274941; CAA95223.1;  
 CC EMBL; U86134; AAB47428.1;  
 CC PIR; JC2068; JC2068.  
 CC SGD; S0005559; DHS1.  
 CC InterPro; IPR000513; Exo\_N\_I.  
 CC InterPro; IPR003584; HHH\_2.

DR InterPro; IPR001532; XPGC\_Rad.  
 DR Pfam; PF00752; XPG\_N; 1.  
 DR Pfam; PF00867; XPG\_I; 1.  
 DR PRINTS; PR00853; XPGRADSUPER.  
 DR SMART; SM00279; Hh2; 1.  
 DR SMART; SM00484; XPGI; 1.  
 DR SMART; SM00485; XPGN; 1.  
 DR PROSITE; PS00841; XPG\_1; 1.  
 DR PROSITE; PS00842; XPG\_2; 1.  
 KW DNA repair; Hydrolase; Nuclease; Endonuclease; Nuclear protein.  
 FT DOMAIN 1 96 N-DOMAIN.  
 FT DOMAIN 114 247 I-DOMAIN.  
 FT DOMAIN 501 520 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 537 553 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 618 625 ASP-RICH (ACIDIC).  
 SQ SEQUENCE 702 AA; 80162 MW; 78E22F6B265DB3AA CRC64;

Query Match 11.2%; Score 218; DB 1; Length 702;  
 Best Local Similarity 23.7%; Pred. No. 5e-07;  
 Matches 99; Conservative 70; Mismatches 176; Indels 72; Gaps 14;

QY 1 MGKIG-LTKLLADNAPKAKSEKESYFGKIAVDASMSIYO-----FLIVVGRGTMET 53  
 Db 1 MGIGLLPQLKPIQNPVSLRR-----YEVEVLAIDGYAWLHRAACSCAYELAMGK----- 50  
 QY 54 LTNEAGEVTSHQGFNTITLLEAGIKPVYVDGKPPDMKKQELAKRYSKRDDATKDLT 113  
 Db 51 -----PTDKYLQFFIKRFSLLTKTFKVEPYLVDFDGAIPVKSTESKRKRKRRE-NKAIA 103  
 QY 114 EAV-EVGDKAIEKLSKRTVKVTRQHNEDCKRLRLMVGVPVVEAPSEAEACALCINDK 172  
 Db 104 ERWAGGEKKNAMYFQRCVDITPEMAKCIYCKLNGIRYIVAFPEADSQMVLQKNI 163  
 QY 173 VFVASEDMDSLTFGAPRFLRLMDPSS-----KKIPVMEFDVAKYLEEL 219  
 Db 164 VOGLISEDSLLVFGCRLLITKLDYGECEICRDNFIKLPKKPLG-----SL 212  
 QY 220 TMOQFDILCIGCDYCDISIKGIGGOTALKLIRHGSIESILENLNKD-RYOIPEDW--P 276  
 Db 213 TNEEITWCLSSCDYTGNGIPKVLITAMKLVRRPNTIERILSIQREGKLMIPDYINE 272  
 QY 277 YQARRLFEPNV-----TLDIPELKWTAPEGLISFLYKDNQNEEDRVTKAEIK 329  
 Db 273 YEAVLAFOQVFCPIRKKIYSLNEIPLYLKDTESKRRRLYACIGFVIHRETQKQIVH 332  
 QY 330 SAKN-----KSSOGRLE--SPFKPTATTSAPLKRKETS-----KTSKAAANKTK 373  
 Db 333 FDDIDHHLHLKIAQDGLNPYDFHOPLANREHLQLASKSNTEFGKTTNTTNSAZVK 389

RESULT 12  
 ID DIN7\_YEAST STANDARD; PRT; 430 AA.  
 AC Q12036;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA-damage inducible protein DIN7.  
 GN DIN7 OR DIN3 OR YDR263C OR YD9320B.02C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OLL;  
 RA MEDLINE=97234602; PubMed=9079876;  
 RA Mieczkowski P.A., Fikus M., Ciesla Z.;  
 RT "Characterization of a novel DNA damage-inducible gene of  
 Saccharomyces cerevisiae, DIN7, which is a structural homolog of the  
 RAD2 and RAD27 DNA repair genes.";  
 RL Mol. Gen. Genet. 253:655-665(1997).  
 RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN-S288c / AB972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: 5'->3' double-stranded DNA exonuclease (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (potential).
CC -!- INDUCTION: By UV light, methyl methane-sulfonate (MMS) or
CC hydroxyurea (HU), and during meiosis.
CC -!- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: X90707; CAA62233.1; -
CC EMBL: 270202; CAA94102.1; -
CC EMBL: 268290; CAA92581.1; -
CC SGD: S0003671; D1N7.
CC DR InterPro; IPR000513; Exo_N_I.
CC DR InterPro; IPR003584; HHH_2.
CC DR InterPro; IPR001532; XPGC_Rad.
CC DR Pfam; PF00752; XPG_N; 1.
CC DR Pfam; PF00867; XPG_I; 1.
CC DR PRINTS; PR00853; XPGRADSUPER.
CC DR SMART; SM00279; HhH2; 1.
CC DR SMART; SM00484; XPGN; 1.
CC DR PROSITE; PS00841; XPG_1; 1.
CC DR PROSITE; PS00842; XPG_2; 1.
CC DR DNA repair; Hydrolase; Nuclease; Endonuclease; Nuclear protein.
FT DOMAIN 114 247 N-DOMAIN.
FT DOMAIN 114 247 I-DOMAIN.
SQ SEQUENCE 430 AA; 49034 MR; 2BC23D30932C79E9 CRC64;

Query Match 10.5%; Score 204.5; DB 1; Length 430;
Best Local Similarity 24.7%; Pred. No. 2.1e-06;
Matches 68; Conservative 53; Mismatches 129; Indels 25; Gaps 6;

QY 1 MGIGKLTLLADNAPKAMKEQKESYFGRKIVADSMISYQ-----FLIVGRTGMETL 54
Db 1 MGIPGLLPQL-----KRIQKQVSLKKYMYQTLAIDGYAWLHRASCACAFELVMNKP----- 51
QY 55 TNEAGEVTSHLQGMFNRTIRLEAGIKPVVYFGKPPDMKKQELAKRYSKRDDATKDLTE 114
Db 52 TNK-----YLQFFIKRLQLRLKIKPKYIVFGDGLFVKNHTETRRKKLENEIMAKK 105
QY 115 AVEVGDKDAIEKLSKRTVKVTRHNECDKRLRLMGVPVVEAPSEAEACALCINDKVF 174
Db 106 LWSAGRNRYNMEYFQKSYVDIIPEMAKIYDYCKLHSIPIYVAPFADPQWYLEKMGILQ 165
QY 175 AVASEMDSLFTGAPRLHMLDPSSKIPVMEFVAKVLEEL---ELTMDQFDILCLIC 231
Db 166 GIISEDSDLLVFGCKTLITLND-QGKALEISKDFSLPENFFILGSEQGFNLVCLA 224
QY 232 GCDYCDISIKGGTALKLRHQHSIESILENLK 266
Db 225 GCDYTSIGIKVGVVTVANKIVKRYSEMKDILIQIER 259

RESULT 13
RA13_SCHPO
ID RA13_SCHPO STANDARD; PRT; 1112 AA.
AC P28706; OS9728;
DT 01-DEC-1992 (Rel. 24, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA repair protein rad13.
GN RAD13 OR SPB3E7.08C.
OS Schizosaccharomyces pombe (fission yeast).

```

```

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC NCBI_TaxID=4896;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=93219111; PubMed=8464724;
RX Carr A.M., Sheldrick K.S., Murray J.M., Al-Harithy R., Watts F.Z.,
RA Lehmann A.R.;
RT "Evolutionary conservation of excision repair in Schizosaccharomyces
RT pombe: evidence for a family of sequences related to the
RT Saccharomyces cerevisiae RAD2 gene.";
RN Nucleic Acids Res. 21:1345-1349(1993).
RN [2]
RX SEQUENCE FROM N.A.
RP MEDLINE=972;
RX Wood V., Swilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Rayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Woestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lenrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: SINGLE-STRANDED DNA ENDONUCLEASE INVOLVED IN EXCISION
CC REPAIR OF DNA DAMAGED WITH UV LIGHT, BULKY ADDUCTS, OR CROSS-
CC LINKING AGENTS. ESSENTIAL FOR THE INCISION STEP OF EXCISION-
CC REPAIR (PROBABLE).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. XPG
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: X66795; CAA47291.1; -
CC DR EMBL; AL023534; CAA19011.1; -
CC DR PIR; S22862; S22862.
CC DR PIR; S30301; S30301.
CC DR InterPro; IPR000513; Exo_N_I.
CC DR InterPro; IPR003584; HHH_2.
CC DR InterPro; IPR003903; UIM.
CC DR InterPro; IPR001044; XPGC_DNA_repair.
CC DR InterPro; IPR001532; XPGC_Rad.
CC DR Pfam; PF00752; XPG_N; 1.
CC DR Pfam; PF00867; XPG_I; 1.
CC DR PRINTS; PR00853; XPGRADSUPER.
CC DR SMART; SM00279; HhH2; 1.
CC DR SMART; SM00484; XPGI; 1.

```

```

DR SMART; SM00485; XPGN; 1.
DR TIGRFAMS; TIGR00600; rad2; 1.
DR PROSITE; PS00841; XPG_1; 1.
DR PROSITE; PS00842; XPG_2; 1.
DR KW DNA repair; Nuclear protein; Hydrolase; Nuclease; Endonuclease.
FT DOMAIN 1 95
FT DOMAIN 742 870 I-DOMAIN.
FT CONFLICT 8 8 D -> N (IN REF. 1).
FT CONFLICT 738 743 LKNOKR -> AOKSKG (IN REF. 1).
FT CONFLICT 738 743 LKNOKR -> AOKSKG (IN REF. 1).
SQ SEQUENCE 1112 AA; 126328 MW; 7ECF4229D5B4768 CRC64;

Query Match 10.5%; Score 204.5; DB 1; Length 1112;
Best Local Similarity 21.4%; Pred. No. 6.7e-06;
Matches 88; Conservative 68; Mismatches 148; Indels 107; Gaps 15;

QY 18 MKEQKESYGRKIV--DASMSIYQFLIVVGRGTMETLNEAGEVTS-----63
Db 604 MKEQADDGTVTNPLNVSSDKMSVY---LLSENAKDT-----GDIKSESIDAVLPTLET 655
QY 64 -----HLOGMNFRTIRLEAGIKPVYFDGKPPDKKOE---LAKRYSKRDDA 108
Db 656 SSPSLISIPDQKEASPNKGAALSSKVEPEV--EKLDEEEEMIIEMAEEREYDRF 713
QY 109 TKDLTE--AVEVGDKDAIEKL-----SKRTVK-----VTRQHNEDCKRLRLMGVPVVE 155
Db 714 VSELNORHETEEWNAFAEKRLKLNQKREKRDADDEVTVQVMIKECQELLRLFLPLVIV 773
QY 156 APSAEAEACALCINDKVFVASEMDSLTFCAPRFLR-----HMDPSSKKIP 204
Db 774 APOEAEACSKLELKVGDIVTDSDFVLEGGTRVYRNMFNQKVFELYLMDMKREFN 833
QY 205 VMEFVAKVLEELNWDQFIDLCILCGDYCDISKIGGQTALKIRO-----253
Db 834 VNQMDLIK-----LAHLGSDVTMGLSRVGVPLAELIHFEPDGTGLFFPK 879
QY 254 -----HGSIET-----ILENLK--DRYOIPEDMPYQEARLKFEPNVLTDIPELKW 298
Db 880 KWFQRLSTGHASKNDVNTVPKRIKLVKGIILFSEFPNPLVDPEALVHPADVDSQSFQW 939
QY 299 TAPDEGLISPLVKNDFNDRVTRKAIEKISAKN-----SSQGRLESFFK 345
Db 940 GIPDLDELQPLMATVGNWSKQRTNEVLLPVIQDMHKKOFVGTQSNLTQFFE 990

RESULT 14
XPG_XENIA
ID XPG_XENIA STANDARD; PRT; 1196 AA.
AC P14629;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA-repair protein complementing XP-G cells homolog (Xeroderma
DE pigmentosum group G complementing protein homolog).
GN ERCC5 OR XPG OR XPGC.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93247645; PubMed=8483504;
RA Scherly D., Nussipikel T., Corlet J., Ucla C., Bairoch A.,
RA Clarkson S.G.;
RT "Complementation of the DNA repair defect in Xeroderma pigmentosum
RL Nature 363:182-185(1993).
CC -!- FUNCTION: SINGLE-STRANDED DNA ENDONUCLEASE INVOLVED IN DNA
CC EXCISION REPAIR. MAKES THE 3' INCISION IN DNA NUCLEOTIDE EXCISION
CC REPAIR (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. XPG
CC SUBFAMILY.

```

```

CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69977; CAA49597.1; -.
DR PIR; S35994; S35994.
DR InterPro; IPR000513; EXO_N.I.
DR InterPro; IPR003584; HHR_2.
DR InterPro; IPR001044; XPGC_DNA_repair.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF00752; XPG_N; 1.
DR Pfam; PF00867; XPG_L; 1.
DR PRINTS; PR00853; XPGRADSUPER.
DR SMART; SM00279; HhH2; 1.
DR SMART; SM00484; XPGI; 1.
DR SMART; SM00485; XPGN; 1.
DR TIGRFAMS; TIGR00600; rad2; 1.
DR PROSITE; PS00841; XPG_1; 1.
DR PROSITE; PS00842; XPG_2; 1.
DR DNA repair; DNA-binding; Nuclear protein; Hydrolase; Nuclease;
DR Endonuclease.
FT DOMAIN 1 95 N-DOMAIN.
FT DOMAIN 786 914 I-DOMAIN.
FT DOMAIN 936 952 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1079 1095 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 1196 AA; 134206 MW; 1FCE1891A3C0623 CRC64;

Query Match 10.3%; Score 200; DB 1; Length 1196;
Best Local Similarity 21.9%; Pred. No. 1.4e-05;
Matches 72; Conservative 64; Mismatches 137; Indels 56; Gaps 11;

QY 88 GPPDMKQELAKRYSKRDDATKDLTEAVEVGDKDAIEK---LSKRTVKYTRHNE---140
Db 740 GEPHEGSNESEK---DLDVDSNQD--ISVELESLENNLYVQVTSLOAQOQQQRIAA 795
QY 141 -----DKRLRLMGVPVVEAPSEAEACALCINDKVFVASEMDSLTFCAPRFL 192
Db 796 TVTQMCLESQELLQFGIPYIVAPMEAEACALDLDTDQTSITDSDSLWFLGARHYV 855
QY 193 RHLMDPSSKKIPVMEFVAKVLEELTMDQFIDLCILCGDYCDISKIGGQTALKIR 252
Db 856 KNEF---SONKHVEYQYADINHQLGDRSKLINLAYLIGSDYTEGPTGVYVSAMEILN 912
QY 253 QH--GSIESILE-----NLNKRYQIP-----EDMPYQEARLKFEP 287
Db 913 EFGQGLEPLVFKKEWSEAKQKKMRPNPNDTKVKKLRLLDLDQOQFPNPVAVASAYLKP 972
QY 288 NVTLDIPELKWTPADEGLISFLVKNDF---NEDRVTKATEKISAKNKSOGRLSEFF 344
Db 973 VPDSEKSAFSGWRPDLEIQIFECESRFGWYRLKTDVLLPV--LKQLNAQQTQLRIDSF 1030
QY 345 KPTATTSAPLKRKETSDKTSKAANKTK 373
Db 1031 RLEQHEAAGLK----SORLRRAVTCMKRK 1055

RESULT 15
DPOOL_THECA
ID DPOOL_THECA STANDARD; PRT; 834 AA.
AC P80194;
DT 01-JUL-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase I, thermostable (EC 2.7.7.7) (TAC polymerase 1).
GN POLA.
OS Thermus caldophilus.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Thermales;
OC Thermaceae; Thermus.

```





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:34:28 ; Search time 58 Seconds  
(without alignments)  
1346.411 Million cell updates/sec

Title: US-09-805-311-2

Perfect score: 1939

Sequence: 1 MGIGLTKLLADNAPKAMKE.....SDKTSKAAANKTKAGGKKK 379

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1807.5	93.2	380	10 Q9SXQ6	Q9sxq6 oryza sativ
2	1313.5	67.7	362	10 Q65231	Q65231 arabidopsis
3	1069	55.1	382	13 P70040	P70040 xenopus lae
4	1068	55.1	382	13 Q57351	Q57351 xenopus lae
5	1068	55.1	382	13 P70054	P70054 xenopus lae
6	1063.5	54.8	381	13 Q90YB0	Q90YB0 gallus gall
7	1043.5	53.8	380	11 Q91Z50	Q91Z50 mus musculus
8	1042	53.7	380	11 Q8R069	Q8R069 mus musculus
9	1040	53.6	380	11 Q9JHW7	Q9JHW7 rattus norv
10	1036.5	53.5	382	5 Q9N3T2	Q9n3t2 caenorhabdi
11	1009.5	52.0	378	11 Q99M86	Q99m86 mus musculus
12	998.5	51.5	385	5 Q96830	Q96830 drosophila
13	887	45.7	650	5 Q9GZ01	Q9gz01 plasmodium
14	887	45.7	672	5 Q9U0K1	Q9u0k1 plasmodium
15	841.5	43.4	345	5 Q8SS91	Q8ss91 encephalito
16	648.5	33.4	340	17 Q93634	Q93634 pyrococcus

17	635	32.8	343	17 Q9V0P9	Q9v0p9 pyrococcus
18	632.5	32.6	343	17 Q50123	Q50123 pyrococcus
19	628	32.4	346	17 Q8ZIN2	Q8zyn2 pyrobaculum
20	622	32.1	348	17 Q8TXU4	Q8txu4 methanopyru
21	586.5	30.2	338	17 Q8TIY5	Q8tiy5 methanosarc
22	542	28.0	328	17 Q27670	Q27670 methanobact
23	538.5	27.8	336	17 Q29975	Q29975 archaeoglob
24	519	26.8	304	17 Q976H6	Q976h6 sulfolobus
25	514.5	26.5	401	17 Q9VEY5	Q9vfy5 aeropyrum p
26	506	26.1	302	17 Q980U8	Q980u8 sulfolobus
27	468	24.1	336	17 Q9HJDA	Q9hjd4 thermoplasma
28	446	23.0	335	17 Q97B98	Q97b98 thermoplasma
29	369	19.0	327	17 Q9HQ27	Q9hq27 halobacteri
30	336	17.3	1516	5 Q96154	Q95154 plasmodium
31	278	14.3	415	12 Q91FF5	Q91ff5 chilo iride
32	268	13.8	366	5 Q8SQX4	Q8sqx4 encephalito
33	267.5	13.8	317	10 Q9C7N8	Q9c7n8 arabidopsis
34	257	13.3	726	5 Q9U9Q6	Q9u9q6 drosophila
35	257	13.3	726	5 Q9VRJ0	Q9vrj0 drosophila
36	244.5	12.6	1037	3 Q8WZX1	Q8wzx1 neurospora
37	243.5	12.6	1257	5 Q917N6	Q917n6 drosophila
38	243	12.5	298	12 Q8QUT3	Q8qut3 infectious
39	241	12.4	298	12 Q9QTF3	Q9qtf3 red sea bre
40	240	12.4	734	13 Q9W6K2	Q9w6k2 xenopus lae
41	235	12.1	497	10 Q9LPD2	Q9lpd2 arabidopsis
42	230	11.9	837	11 Q9QZ11	Q9qz11 mus musculu
43	230	11.9	837	11 Q9Z3A5	Q9z3a5 mus musculu
44	229	11.8	390	12 Q8QY21	Q8qy21 rana tigrin
45	227	11.7	800	4 Q75466	Q75466 homo sapien

## ALIGNMENTS

### RESULT 1

Q9SXQ6 Q9SXQ6 PRELIMINARY; PRT; 380 AA.  
ID Q9SXQ6; AC Q9SXQ6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE FEH-1  
GN OSFEN-1.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20256470; PubMed=10798612;  
RA Kimura S., Ueda T., Hatanaka M., Takenouchi M., Hashimoto J.,  
RA Sakaguchi K.;  
RT "plant homologue of flap endonuclease-1: molecular cloning,  
RT characterization, and evidence of expression in meristematic  
RT tissues.";  
RL Plant Mol. Biol. 42:415-427(2000)  
DR EMBL; AB021666; BAA36171.1; -  
DR HSP; Q58839; IA76.  
DR InterPro; IPR002421; 5\_3\_exonuclease.  
DR InterPro; IPR000513; Exo\_N\_I.  
DR InterPro; IPR003584; HHH\_2.  
DR InterPro; IPR001532; XPGC\_Rad.  
DR Pfam; PF01367; 5\_3\_exonuclease; 1.  
DR Pfam; PF00867; XPG\_I; 1.  
DR Pfam; PF00752; XPG\_N; 1.  
DR PRINTS; PR00853; XPGRADSUPER.  
DR SMART; SM00475; 53EXOC; 1.  
DR SMART; SM00279; Hhr2; 1.  
DR SMART; SM00484; XPGI; 1.  
DR SMART; SM00485; XPGN; 1.  
DR PROSITE; PS00841; XPG\_1; 1.  
SQ SEQUENCE 380 AA; 42792 MW; E0148AAFA95A7364 CRC64;

```
Query Match          93.2%; Score 1807.5; DB 10; Length 380;
Best Local Similarity 92.1%; Pred. No. 9.8e-120;
Matches 350; Conservative 16; Mismatches 13; Indels 1; Gaps 1;

QY 1 MGKIGLTKLLADNAPKAMEQKESYFGKRIAVDASMSIYQFLIVVGRGTGNETLTNEAGE 60
DB 1 MGKIGLTKLLADNAPKAMEQKESYFGKRIAVDASMSIYQFLIVVGRGTGNETLTNEAGE 60
QY 61 VTSHLQGMFNRTIRLLEAGIKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTAEVGD 120
DB 61 VTSHLQGMFNRTIRLLEAGIKPVYVDFGKPPDLKKQELAKRYSKREDATKELTAEVEGD 120
QY 121 KDAIEKLSKRTYKVTQHNEDCKRLRLRLMGVPVVEAPSEAECAALCINDKVFVASED 180
DB 121 KDAIEKSPKRTYKVTQHNEDCKRLRLRLMGVPVVEAPSEAECAALCINDKVFVASED 180
QY 181 MDSLTFGAPRFLHLMDSPPSKKIPVMEFDVAKVLELELTMDQFDILCGCDYCDISIK 240
DB 181 MDSLTFGAPRFLHLMDSPPSKKIPVMEFEVAKVLELELTMDQFDILCGCDYCDISIK 240
QY 241 GIGGQATALKLIRHOGSIESILENLKRDYQIPEDWPYQEARLKEPNVTLDIPELKWA 300
DB 241 GIGGQATALKLIRHOGSIESILENLKRDYQIPEDWPYQEARLKEPNVTLDIPELKWA 300
QY 301 PDEGLISFLVNDGNGFNEDRVTKAIEKTKSAKNKSSQGRLESFFKPTATTSAPLKRKETS 360
DB 301 PDEGLVFLVKNFGNDRVTKAIEKTKFAKNKSSQGRLESFFKPVVYSTVPLKRKETS 360
QY 361 DKTSKAAANKTK-AGGKKK 379
DB 361 EKPTKAVANKTKGAGGKKK 380

RESULT 2
065251 PRELIMINARY; PRT; 362 AA.
AC 065251;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE F21E10.3 protein.
GN F21E10.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Davidson S., Rohlfing T., David M., O'Brian D.;
RT "The sequence of A. thaliana F21E10."
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Washu;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Wilson R.;
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF058914; AAC13596.1;
RL HSSP; Q58839; 1A76.
DR InterPro; IPR002421; 5_3_exonuclease.
DR

Query Match          67.7%; Score 1313.5; DB 10; Length 362;
Best Local Similarity 72.1%; Pred. No. 6.9e-85;
Matches 259; Conservative 30; Mismatches 33; Indels 37; Gaps 3;

QY 1 MGKIGLTKLLADNAPKAMEQKESYFGKRIAVDASMSIYQFLIVVGRGTGNETLTNEAGE 60
DB 1 MGKIGLTKLLADNAPSCMKQKESYFGKRIAVDASMSIYQFLVRAPSSPL-LLEELGL 59
QY 61 VTS-----HLQGMFNRTIRLLEAGIKPVYVF 86
DB 60 KCSLMKLVSLWIFLDCVLCIYVFFVWVYMMFNRLHOGMFRNTRLLEAGIKPVYVF 119
QY 87 DGKPPDMKKQELAKRYSKRDDATKDLTAEVGVGDKAIEKLSKRTYKVTQHNEDCKRL 146
DB 120 DGKPPELKQELAKRYSKRADATADLTGAEAGNKEDIEKYSKRTYKVTQHNEDCKRL 179
QY 147 RLMGVPVVEAPSEAECAALCINDKVFVASEDMDSLTGAPRFLHLMDSPPSKKIPVM 206
DB 180 RLMGVPVVEATSEAECAALCKSGKVGAVASEDMDSLTGAPRFLHLMDSPPSKKIPVM 239
QY 207 EFDVAKVLELELTMDQFDILCGCDYCDISIRGIGGQATALKLIRHOGSIESILENLK 266
DB 240 EFEVAKILEELQTLTMDQFDILCGCDYCDISIRGIGGQATALKLIRHOGSIESILENLK 299
QY 267 DRYQIPEDWPYQEARLKEPNVTLDIPEL--KWTAPDEGLISFLVKNKGNEFDRVTK 323
DB 300 ERYQIPPEWPYNEARKLFKEPVDITDESQDLTKWTSPEDEGIVQFLVNGFNDRVTK 358

RESULT 3
P70040 PRELIMINARY; PRT; 382 AA.
AC P70040;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE 5' nuclease XFENIA.
GN FEN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Piploidea;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RA Bibikova M., Chi E., Wu B., Kim K., Carroll D.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Li J.-L., Cox L.S.;
RT "Cloning and investigation of Xenopus FEN1: developmental expression
and function in DNA replication."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U64563; AAB06176.1;
DR EMBL; AF065397; AAD02814.1;
DR HSSP; Q58839; 1A76.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR000513; Exo_N.1.
DR InterPro; IPR003584; HHH_2.
DR InterPro; IPR001532; XPGC_Rad.
```



```
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00867; XPG_I; 1.
DR Pfam; PF00752; XPG_N; 1.
DR PRINTS; PR00853; XPGRAUSUPER.
DR SMART; SM00279; HH2; 1.
DR SMART; SM00484; XPGI; 1.
DR SMART; SM00485; XPGN; 1.
DR Endonuclease.
KW Endonuclease.
SQ SEQUENCE 382 AA; 42682 MW; 9B1DB0EDAD158D57 CRC64;

Query Match
Best Local Similarity 55.1%; Score 1069; DB 13; Length 382;
Matches 210; Conservative 63; Mismatches 98; Indels 6; Gaps 4;

QY 1 MGTKGLTKLLADNAPKAMKEQKESYFGKRIYAVDASMSIYQFLIVVGRGTGNETLTNEAGE 60
DB 1 MGHLGLAKLIADVAPAAIKHEHDIKSYFGKRVAVDASMCIIYQFLIAYRQDG-NTLQNEEGE 59

QY 61 VTSHLOGMFNRTIRLLEAGIKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVG 120
DB 60 TTSMLMGMYRTIRVHEGIRKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVG 119

QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVVPVVEAPSEAEAEACALCINDKVFVASED 180
DB 120 VENIEKTRKLVKVTQHNEDCKRLRLMGVVPVVEAPSEAEAEACALCINDKVFVASED 179

QY 181 MSLTFGAPRFLRLHMDSPSKIPVMEFDVAKVLELELTMDQFIDLCILCGCDYCSIK 240
DB 180 MDALTFGTPVLLRHULTASEAKKLPIQEFHLNRVQDITGITHQFVDFVLCILGSDYCTIR 239

QY 241 GIGGQATLKLIROHGSIESILENKNDRYQIPEDPWYQEARRLFKEPNVT-LDPELKW 299
DB 240 GIGPKRAIDLIRQHTIDEIINDLKYVPVPEWNLHKEAKHLFLEPEVVDITELKWI 299

QY 300 APDEGLISFLVKNDFNEDRVTKAEIKSAKNKSOGLESEFFKPTATTSAPLRKET 359
DB 300 EPDEGLVAFNGCEKQFSEDRIRNGAKLAKNRQSTQGRDLDFFKVTGSVSS-TKRKEA 358

QY 360 SDKTSKAAANKTKAGG 376
DB 359 E---SKGSAKKAKTGG 372

RESULT 4
O57351 PRELIMINARY; PRT; 382 AA.
AC O57351
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Flap endonuclease 1.
GN FEN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98204872; PubMed=9535864;
RA Kim K., Blade S., Matsumoto Y.;
RT "Involvement of flap endonuclease 1 in base excision DNA repair.";
RL J. Biol. Chem. 273:8842-8848(1998).
DR EMBL; AF036327; AAB88707.1;
DR HSSP; Q58839; 1A76.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR000513; Exo_N1.
DR InterPro; IPR003584; HHH_2.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00867; XPG_I; 1.
DR Pfam; PF00752; XPG_N; 1.
DR PRINTS; PR00853; XPGRAUSUPER.
DR SMART; SM00484; XPGI; 1.
DR SMART; SM00485; XPGN; 1.

QY 1 MGTKGLTKLLADNAPKAMKEQKESYFGKRIYAVDASMSIYQFLIVVGRGTGNETLTNEAGE 60
DB 1 MGHLGLAKLIADVAPAAIKHEHDIKSYFGKRVAVDASMCIIYQFLIAYRQDG-NTLQNEEGE 59

QY 61 VTSHLOGMFNRTIRLLEAGIKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVG 120
DB 60 TTSMLMGMYRTIRVHEGIRKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVG 119

QY 121 KDAIEKLSKRTVKVTRQHNEDCKRLRLMGVVPVVEAPSEAEAEACALCINDKVFVASED 180
DB 120 VENIEKTRKLVKVTQHNEDCKRLRLMGVVPVVEAPSEAEAEACALCINDKVFVASED 179

QY 181 MSLTFGAPRFLRLHMDSPSKIPVMEFDVAKVLELELTMDQFIDLCILCGCDYCSIK 240
DB 180 MDALTFGTPVLLRHULTASEAKKLPIQEFHLNRVQDITGITHQFVDFVLCILGSDYCTIR 239

QY 241 GIGGQATLKLIROHGSIESILENKNDRYQIPEDPWYQEARRLFKEPNVT-LDPELKW 299
DB 240 GIGPKRAIDLIRQHTIDEIINDLKYVPVPEWNLHKEAKHLFLEPEVVDITELKWI 299

QY 300 APDEGLISFLVKNDFNEDRVTKAEIKSAKNKSOGLESEFFKPTATTSAPLRKET 359
DB 300 EPDEGLVAFNGCEKQFSEDRIRNGAKLAKNRQSTQGRDLDFFKVTGSVSS-TKRKEA 358

QY 360 SDKTSKAAANKTKAGG 376
DB 359 E---SKGSAKKAKTGG 372

RESULT 5
P70054 PRELIMINARY; PRT; 382 AA.
AC P70054
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE XEN1b.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069415; PubMed=9852084;
RA Bibikova M., Wu B., Chi E., Kim K.H., Trautman J.K., Carroll D.;
RT "Characterization of FEN-1 from Xenopus laevis. cDNA cloning and role
in DNA metabolism.";
RL J. Biol. Chem. 273:34222-34229(1998).
DR EMBL; U68141; AAB08478.1;
DR HSSP; Q58839; 1A76.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR000513; Exo_N1.
DR InterPro; IPR003584; HHH_2.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00867; XPG_I; 1.
DR Pfam; PF00752; XPG_N; 1.
DR PRINTS; PR00853; XPGRAUSUPER.
DR SMART; SM00279; HH2; 1.
DR SMART; SM00484; XPGI; 1.
DR SMART; SM00485; XPGN; 1.
```

```

SQ SEQUENCE 382 AA; 42865 MW; 1C648936A232D460 CRC64;
Query Match 55.1%; Score 1068; DB 13; Length 382;
Best Local Similarity 56.2%; Pred. No. 1.6e-67;
Matches 212; Conservative 60; Mismatches 99; Indels 6; Gaps 4;

QY 1 MGIGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIIVGRTGMETLTNEAGE 60
DB 1 MGIGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIIVGRTGMETLTNEAGE 59
QY 61 VTSHLQGMFNRTIRLLEAGIKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120
DB 60 TTSHLGMGFYRTIRLLEAGIKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 119
QY 121 KDAIEKLSKRTVKTROHNECDCKRLRLMGVVPVEAPSEAECAALCINDKVFVASED 180
DB 120 VENIEKFNKRLVKTROHNECDCKRLRLMGVVPVEAPSEAECAALCINDKVFVASED 179
QY 181 MDSLTGAPFLRLHMDPSSKKIPVMEFDVAKVLELELTMDQFIDLCILGCGDYCDSIK 240
DB 180 MDALTFTGTPVLLRLHTASEAKKLPIQEFHLNRVFDQIDGINHEQFVDLCILGSDYCETIR 239
QY 241 GIGGOTALKLIRHOGSTESILENLNDRYOIPEDWPYQEARLFKEPNV-TLIDPELKWT 299
DB 240 GIGPKRAIDLIRHOGSTESILENLNDRYOIPEDWPYQEARLFKEPNV-TLIDPELKWT 299
QY 300 APDEGLISFLVNDGNEFDVRYTKAIEKSKAKNKSQGRLESEFFKPTAT-TSAPLKRKE 358
DB 300 EPDEGLVAFMCGEKQFSEDIRNGAKLAKNRQGSTQGRLLDDFFKVTGTSISS-TKRKEY 358
QY 360 SDKTSKAAANKTKRAGG 376
DB 359 E---SKGTSKKKSTGG 372

RESULT 6
Q90YB0 PRELIMINARY; PRT; 381 AA.
AC Q90YB0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE FEN-1 nuclease.
GN FEN1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsuzaki Y., Adachi N., Koyama H.;
RT "The FEN-1 nuclease is not essential for DNA replication, but is
RT required for maintenance of genomic integrity and base excision repair
RT in vertebrate cells."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB058602; BAB68507.1; -.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR000513; Exo_N_1.
DR InterPro; IPR003583; HHH_1.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF00867; XPG_I; 1.
DR Pfam; PF00752; XPG_N; 1.
DR SMART; SM00278; Hhh1; 1.
DR PROSITE; PS00841; XPG_1; UNKNOWN_1.
DR PROSITE; PS00842; XPG_2; UNKNOWN_1.
KW Phosphopantetheine.
SQ SEQUENCE 381 AA; 43054 MW; 71432ECBCB4E67AA CRC64;
Query Match 54.8%; Score 1063.5; DB 13; Length 381;
Best Local Similarity 54.2%; Pred. No. 3.3e-67;
Matches 206; Conservative 67; Mismatches 104; Indels 3; Gaps 3;

```

```

QY 1 MGIGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIIVGRTGMETLTNEAGE 60
DB 1 MGIGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIIVGRTGMETLTNEAGE 59
QY 61 VTSHLQGMFNRTIRLLEAGIKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120
DB 60 TTSHLGMGFYRTIRLLEAGIKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 119
QY 121 KDAIEKLSKRTVKTROHNECDCKRLRLMGVVPVEAPSEAECAALCINDKVFVASED 180
DB 120 EANTIEKFNKRLVKTROHNECDCKRLRLMGVVPVEAPSEAECAALCINDKVFVASED 179
QY 181 MDSLTGAPFLRLHMDPSSKKIPVMEFDVAKVLELELTMDQFIDLCILGCGDYCDSIK 240
DB 180 MDCLTFGSPVLRHLTASEAKKLPIQEFHLNRVLDLGLTWEQFVDLCILGDCYCESIR 239
QY 241 GIGGOTALKLIRHOGSTESILENLNDRYOIPEDWPYQEARLFKEPNV-TLIDPELKWT 299
DB 240 GIGPKRAIDLIRHOGSTESILENLNDRYOIPEDWPYQEARLFKEPNV-TLIDPELKWT 299
QY 300 APDEGLISFLVNDGNEFDVRYTKAIEKSKAKNKSQGRLESEFFKPTAT-TSAPLKRKE 358
DB 300 EPNEEELVQPMCGEKQFSEDIRNGVVKRLSKSQGSTQGRLLDDFFKVTGTSISAKRKEPE 359
QY 359 TSDKTSKAAANKTKRAGGK 378
DB 360 TKSAAKKTNSATAKFKK 379

RESULT 7
Q91250 PRELIMINARY; PRT; 380 AA.
AC Q91250;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to flag structure-specific endonuclease.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010203; AAH10203.1; -.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR000513; Exo_N_1.
DR InterPro; IPR003583; HHH_1.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00867; XPG_I; 1.
DR Pfam; PF00752; XPG_N; 1.
DR SMART; SM00278; Hhh1; 1.
DR PROSITE; PS00841; XPG_1; UNKNOWN_1.
DR PROSITE; PS00842; XPG_2; UNKNOWN_1.
KW Endonuclease.
SQ SEQUENCE 380 AA; 42623 MW; 1BE903288B46520D CRC64;
Query Match 53.8%; Score 1043; DB 11; Length 380;
Best Local Similarity 53.8%; Pred. No. 9.2e-66;
Matches 205; Conservative 69; Mismatches 99; Indels 8; Gaps 5;

QY 1 MGIGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIIVGRTGMETLTNEAGE 60
DB 1 MGIGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIIVGRTGMETLTNEAGE 59
QY 61 VTSHLQGMFNRTIRLLEAGIKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120
DB 60 TTSHLGMGFYRTIRLLEAGIKPVYVDFGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 119
QY 121 KDAIEKLSKRTVKTROHNECDCKRLRLMGVVPVEAPSEAECAALCINDKVFVASED 180

```

```
Db 120 EEEVEKFTKRLVKYTKQHNDCKHLLSLMGIPYLDAPSSEAEASCAALAKAGKYVAATED 179
Qy 181 MDSLTFGAPRFLRHLMDPSSKKIPVMEFDYAKVLELELTMDQFDLCLCGCYCDSIK 240
Db 180 MDCLTGPGSVLMRHLTASEAKKLPQIEFHLSRVQLGSLNQEQFVLDLCLGSDYCESIR 239
Qy 241 GIGGQFALKLRHQHSIESILENLKNDROYIPEDWPYQEARLFKEPNVTL--IPELKW 298
Db 240 GIGPKRAVDLIQKHKSIEIIVRRLDPSKYPVPENWLKHAQAQFLPEV-LDPESVELKW 298
Qy 299 TAPDEGLISFLVKGNGFNEDRVTKATEIKSAKNSSOGRLSEFPFKPTATTAPLKRKE 358
Db 299 SEPNEELVKFCGKEQFSEERIRSGVKRLSKSQGSTQGRLLDDFFKVTGSLSS-AKRKE 357
Qy 359 TSDKTSKAAANKTKAGGKK 379
Db 358 PE---PKGPAKKAKTGAGK 375

RESULT 8
Q8R069 PRELIMINARY; PRT; 380 AA.
AC Q8R069;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE Hypothetical 42.6 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
BL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027295; AAF27295.1;
KW Hypothetical protein.
SQ SEQUENCE 380 AA; 42609 MW; C8B38745358F65AF CRC64;

Query Match 53.7%; Score 1042; DB 11; Length 380;
Best Local Similarity 53.5%; Pred. No. 1.1e-65;
Matches 204; Conservative 70; Mismatches 99; Indels 8; Gaps 5;

Qy 1 MGIKGLTKLLADNAPKAMKEQKESYFGKRIAYDASMSIYQFLIIVGRTGMETLTNEAGE 60
Db 1 MGTHGLAKLIADVAPSARENDIKSYFGKRAVAIDASMSIYQFLIIV-RQGQGVQLQNEEGE 59

Qy 61 VTSHLOGMFNRTIRLEAGIKPVYFDGKPPDMKQELAKRYSKRDDATKDLTEAVEVGD 120
Db 60 TTSHLMGMYRTIRMENGAVPVYFDGKPPQLKSGELASERRAEAKLOQAQAGM 119

Qy 121 KDAIEKLSRTVKYTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFAYASED 180
Db 120 EEEVEKFTKRLVKYTKQHNDCKHLLSLMGIPYLDAPSSEAEASCAALAKAGKYVAATED 179

Qy 181 MDSLTFGAPRFLRHLMDPSSKKIPVMEFDYAKVLELELTMDQFDLCLCGCYCDSIK 240
Db 180 MDCLTGPGSVLMRHLTASEAKKLPQIEFHLSRVQLGSLNQEQFVLDLCLGSDYCESIR 239

Qy 241 GIGGQFALKLRHQHSIESILENLKNDROYIPEDWPYQEARLFKEPNVTL--IPELKW 298
Db 240 GIGPKRAVDLIQKHKSIEIIVRRLDPSKYPVPENWLKHAQAQFLPEV-LDPESVELKW 298

Qy 299 TAPDEGLISFLVKGNGFNEDRVTKATEIKSAKNSSOGRLSEFPFKPTATTAPLKRKE 358
Db 299 SEPNEELVKFCGKEQFSEERIRSGVKRLSKSQGSTQGRLLDDFFKVTGSLSS-AKRKE 357

Qy 359 TSDKTSKAAANKTKAGGKK 379
Db 358 PE---PKGPAKKAKTGAGK 375
```

```
RESULT 9
Q9JHW7 PRELIMINARY; PRT; 380 AA.
ID Q9JHW7;
AC Q9JHW7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Flag structure-specific endonuclease.
GN Fen-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen D., Cao G., Yang S., Li M., Chen J.;
RT "Cloning and characterization of a rat DNA structure-specific
RT endonuclease (Fen-1).", the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF281018; AAF81265.1;
DR HSSP; Q58639; 1A76.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR000513; Exo_N1.
DR InterPro; IPR003584; HHH_2.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00867; XPG_I; 1.
DR Pfam; PF00752; XPG_N; 1.
DR PRINTS; PR00853; XPGRADSUPER.
DR SMART; SM00279; Hhh2; 1.
DR SMART; SM00484; XPGT; 1.
DR SMART; SM00485; XPGN; 1.
DR PROSITE; PS00842; XPG_2; 1.
DR Endonuclease.
KW SEQUENCE 380 AA; 42622 MW; 61698CE16F182136 CRC64;

Query Match 53.6%; Score 1040; DB 11; Length 380;
Best Local Similarity 52.6%; Pred. No. 1.5e-65;
Matches 200; Conservative 73; Mismatches 101; Indels 6; Gaps 4;

Qy 1 MGIKGLTKLLADNAPKAMKEQKESYFGKRIAYDASMSIYQFLIIVGRTGMETLTNEAGE 60
Db 1 MGTHGLAKLIADVAPSARENDIKSYFGKRAVAIDASMSIYQFLIIV-RQGQGVQLQNEEGE 59

Qy 61 VTSHLOGMFNRTIRLEAGIKPVYFDGKPPDMKQELAKRYSKRDDATKDLTEAVEVGD 120
Db 60 TTSHLMGMYRTIRMENGAVPVYFDGKPPQLKSAQLAKRSERRAEAKLOQAQAGM 119

Qy 121 KDAIEKLSRTVKYTRQHNEDCKRLRLMGVPVVEAPSEAEACALCINDKVFAYASED 180
Db 120 EEEVEKFTKRLVKYTKQHNDCKHLLSLMGIPYLDAPSSEAEASCAALAKAGKYVAATED 179

Qy 181 MDSLTFGAPRFLRHLMDPSSKKIPVMEFDYAKVLELELTMDQFDLCLCGCYCDSIK 240
Db 180 MDCLTGPGSVLMRHLTASEAKKLPQIEFHLSRVQLGSLNQEQFVLDLCLGSDYCESIR 239

Qy 241 GIGGQFALKLRHQHSIESILENLKNDROYIPEDWPYQEARLFKEPNVTL--IPELKW 299
Db 240 GIGPKRAVDLIQKHKSIEIIVRRLDPSKYPVPENWLKHAQAQFLPEV-LDPESVELKW 299

Qy 300 APDEGLISFLVKGNGFNEDRVTKATEIKSAKNSSOGRLSEFPFKPTATTAPLKRKE 359
Db 300 EPNEELVKFCGKEQFSEERIRSGVKRLSKSQGSTQGRLLDDFFKVTGSLSS-AKRKE 358

Qy 360 SDKTSKAAANKTKAGGKK 379
Db 359 E---PKGPAKKAKTGAGK 375

RESULT 10
Q9N3T2 PRELIMINARY; PRT; 382 AA.
ID Q9N3T2;
AC Q9N3T2;
```

```

DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical 42.5 kDa protein.
GN Y47G6A.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
None;
RA "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Minx P., Graves T., Hawrysko C.;
RT "The sequence of C. elegans cosmid Y47G6A.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024791; AAF60853.1;
DR HSSP; Q58839; IA76.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR000513; EXO_N_1.
DR InterPro; IPR003584; HH2.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00867; XPG_I; 1.
DR Pfam; PF00752; XPG_N; 1.
DR PRINTS; PR00853; XPG_RADSUPER.
DR SMART; SM00279; HH2; 1.
DR SMART; SM00484; XPGI; 1.
DR SMART; SM00485; XPGN; 1.
DR PROSITE; PS00841; XPG_1; 1.
DR PROSITE; PS00842; XPG_2; 1.
DR Endonuclease.
KW Hypothetical protein.
SQ SEQUENCE 382 AA; 42549 MW; FDFE0BA0707321EB CRC64;

Query Match 53.5%; Score 1036.5; DB 5; Length 382;
Best Local Similarity 52.3%; Pred. No. 2.7e-65;
Matches 201; Conservative 73; Mismatches 99; Indels 11; Gaps 4;

QY 1 MGKIGLTKLLADNAPKAMKQKESYFGKRIADVADSMISYQFLIVVGTGNETILTNEAGE 60
DB 1 MGKIGLTKLLADNAPKAMKQKESYFGKRIADVADSMISYQFLIVVGTGNETILTNEAGE 60
QY 61 VTSHLQGMFNRTIRLLRAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120
DB 61 VTSHLQGMFNRTIRLLRAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120
QY 60 TTS-SLMGMYFTIR-MENGKIPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 117
DB 60 TTS-SLMGMYFTIR-MENGKIPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 117
QY 121 KDAIEKLSKRVKVTYTRHNDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180
DB 121 KDAIEKLSKRVKVTYTRHNDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180
QY 120 VKEAEKFERRLVKYTKQONDEAKRLGLMGIPVVEAPSEAEACALCINDKVFVASED 179
DB 120 VKEAEKFERRLVKYTKQONDEAKRLGLMGIPVVEAPSEAEACALCINDKVFVASED 179
QY 181 MDSLTFGAPRFLRHLPDPSKKIPVMEFVAVKLELELTMDQFIDLCILCGDYCDSTK 240
DB 181 MDSLTFGAPRFLRHLPDPSKKIPVMEFVAVKLELELTMDQFIDLCILCGDYCDSTK 240
QY 180 MDALTFGSLVLRHFLAPVAKKIPKEFNLSLALAEEMKLSVEEFIDLCILGDCYCTIR 239
DB 180 MDALTFGSLVLRHFLAPVAKKIPKEFNLSLALAEEMKLSVEEFIDLCILGDCYCTIR 239
QY 241 GIGGQTALKLRQHGSIESTILENLKNDKRYQIPEDPMPYQEARLKFEPNVT-LDIPELKWT 299
DB 241 GIGGQTALKLRQHGSIESTILENLKNDKRYQIPEDPMPYQEARLKFEPNVT-LDIPELKWT 299
QY 240 GVGPKKAVELIRQHNKTIETILENDONKYPPEDPMPYKRAELFLNPVTKPPEVELTWK 299
DB 240 GVGPKKAVELIRQHNKTIETILENDONKYPPEDPMPYKRAELFLNPVTKPPEVELTWK 299
QY 300 APDEGLISFLVKGNGFNEDRVTKAIEKISAKNKSQGRLESFF---KPTATTAPLK 355
DB 300 APDEGLISFLVKGNGFNEDRVTKAIEKISAKNKSQGRLESFF---KPTATTAPLK 355
QY 300 EADVEGVIFLQCGEKNFNEIRIALKLKTSKSGTQGRIDSFSGNSTKVTCTVATTKR 359
DB 300 EADVEGVIFLQCGEKNFNEIRIALKLKTSKSGTQGRIDSFSGNSTKVTCTVATTKR 359
```

```

QY 356 RKETSDDKTSKAAANKTKAGGKK 379
DB 360 KAEAEAKAKGA-----KGGPPK 378

RESULT 11
Q99M86 PRELIMINARY; PRT: 378 AA.
AC Q99M86;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Flap endonuclease-1.
GN FEN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV/J;
RA Karanjawala Z.E., Shi X., Hsieh C.-L., Lieber M.R.;
RT "The Mammalian Fen1 Locus: Structure and Conserved Sequence
Features.";
RL Microb. Comp. Genomics 0:0-0(2001).
DR EMBL; AY014962; AA01853.1;
DR HSSP; Q58839; IA76.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR000513; EXO_N_1.
DR InterPro; IPR003584; HH2.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00867; XPG_I; 1.
DR Pfam; PF00752; XPG_N; 1.
DR PRINTS; PR00853; XPG_RADSUPER.
DR SMART; SM00279; HH2; 1.
DR SMART; SM00484; XPGI; 1.
DR SMART; SM00485; XPGN; 1.
DR PROSITE; PS00841; XPG_1; 1.
DR PROSITE; PS00842; XPG_2; 1.
DR Endonuclease.
KW Endonuclease.
SQ SEQUENCE 378 AA; 42300 MW; 827946B8AC9F39 CRC64;

Query Match 52.0%; Score 1009; DB 11; Length 378;
Best Local Similarity 52.9%; Pred. No. 2.3e-63;
Matches 201; Conservative 69; Mismatches 102; Indels 8; Gaps 6;

QY 1 MGKIGLTKLLADNAPKAMKQKESYFGKRIADVADSMISYQFLIVVGTGNETILTNEAGE 60
DB 1 MGKIGLTKLLADNAPKAMKQKESYFGKRIADVADSMISYQFLIVVGTGNETILTNEAGE 60
QY 61 VTSHLQGMFNRTIRLLRAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120
DB 61 VTSHLQGMFNRTIRLLRAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 120
QY 60 TTS-SLMGMYFTIR-MENGKIPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 117
DB 60 TTS-SLMGMYFTIR-MENGKIPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVGD 117
QY 121 KDAIEKLSKRVKVTYTRHNDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180
DB 121 KDAIEKLSKRVKVTYTRHNDCKRLRLMGVPVVEAPSEAEACALCINDKVFVASED 180
QY 118 EBEVEKFTKRLVKVTKQNDCKHLLSLMGIPYLDAPSEAEASCAALAKAKVYAAATED 177
DB 118 EBEVEKFTKRLVKVTKQNDCKHLLSLMGIPYLDAPSEAEASCAALAKAKVYAAATED 177
QY 181 MDSLTFGAPRFLRHLPDPSKKIPVMEFVAVKLELELTMDQFIDLCILCGDYCDSTK 240
DB 181 MDSLTFGAPRFLRHLPDPSKKIPVMEFVAVKLELELTMDQFIDLCILCGDYCDSTK 240
QY 178 MDCLTFGSLVLRHFLAPVAKKIPKEFNLSLALAEEMKLSVEEFIDLCILGDCYCTIR 237
DB 178 MDCLTFGSLVLRHFLAPVAKKIPKEFNLSLALAEEMKLSVEEFIDLCILGDCYCTIR 237
QY 241 GIGGQTALKLRQHGSIESTILENLKNDKRYQIPEDPMPYQEARLKFEPNVT-LDIPELKWT 299
DB 241 GIGGQTALKLRQHGSIESTILENLKNDKRYQIPEDPMPYQEARLKFEPNVT-LDIPELKWT 299
QY 238 GIGAKRVDLIRQHNKTIETILENDONKYPPEDPMPYKRAELFLNPVTKPPEVELKWS 297
DB 238 GIGAKRVDLIRQHNKTIETILENDONKYPPEDPMPYKRAELFLNPVTKPPEVELKWS 297
QY 300 APDEGLISFLVKGNGFNEDRVTKAIEKISAKNKSQGRLESFFKPTATTAPLK 359
DB 300 APDEGLISFLVKGNGFNEDRVTKAIEKISAKNKSQGRLESFFKPTATTAPLK 359
QY 298 ENEEELVKFCGCKQSEERIRSGVKRLSKRQSGTQGRIDDFNVTGSLSS-AKKKEP 356
DB 298 ENEEELVKFCGCKQSEERIRSGVKRLSKRQSGTQGRIDDFNVTGSLSS-AKKKEP 356
QY 360 SDKTSKAAANKTKAGGKK 379
DB 360 SDKTSKAAANKTKAGGKK 379
```

```

Db 357 E---PKGPAKKAKTGACK 373
RESULT 12
O96830
AC O96830 PRELIMINARY; PRT; 385 AA.
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
EG:EG0003.3 protein.
GN FEN1 OR EG:EG0003.3 OR C8648.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agayani A., An H.-J., Andrews-Pfannkoche C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gangel A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M.D., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissentbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.-H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[2]
RN SEQUENCE FROM N.A.
RP Murphy L., Harris D., Barrell B.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster."
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Benos P.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003805; RAF57944.1; -
DR EMBL; AL031863; CAA21320.1; -
DR HSSP; Q58839; 1A76.
DR FlyBase; FBgn0025632; FEN1.
DR InterPro; IPR002421; 5_3_exonuclease.
DR
DR InterPro; IPR000513; Exo_N_I.
DR InterPro; IPR003584; HHH_2.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR InterPro; IPR000513; Exo_N_I.
DR InterPro; IPR003584; HHH_2.
DR Pfam; PF00867; XPG-I; 1.
DR Pfam; PF00752; XPG-N; 1.
DR PRINTS; PR00853; XPGRADSUPER.
DR SMART; SM00278; HHH1; 1.
DR SMART; SM00279; HHH2; 1.
DR SMART; SM00484; XPG1; 1.
DR SMART; SM00485; XPGN; 1.
DR PROSITE; PS00841; XPG.1; 1.
DR PROSITE; PS00842; XPG.2; 1.
SQ SEQUENCE 385 AA; 42948 MW; DOA0831C2BDA9240 CRC64;
Query Match 51.5%; Score 998.5; DB 5; Length 385;
Best Local Similarity 52.9%; Pred. No. 1.3e-62;
Matches 202; Conservative 65; Mismatches 104; Indels 11; Gaps 5;
QY 1 MGIKGLTKLLADNAPKAMKEQKESYFGRKIAYDASMSIQFLIVVGRGTGMETLTNEAGE 60
DB 1 MGILGLSLIADLAPOAIRSESEMKHFFGRKVAIDASMCILYQFLIAYRSEGAQ-LATVNGD 59
QY 61 VTSHLOGNFNTIRLLLEAGIKPVYFDGKPPDKMKQELAKRYKRDATKDLTEAVEVD 120
DB 60 PTVHLMGMFYITIRLLONGIKPVYFDGKPPDLKSGELAKRAREBEAKALKAAATDAGD 119
QY 121 KDAIEKLSKRVKVTROHNECDKRLRLMLGVPVVEAPSEAEACALCINDKVFAYASED 180
DB 120 DAGIEKFNRLRVVTKHAKKELLTLMGVVVDAPCEAEACALVAKGVYATATED 179
QY 181 VDSLTFGAPRLHMDPSSKKIPVMEFDVAKVLEELTMDQFIDLCILCGDYCDISK 240
DB 180 MDALFTGSKLRITVSEAKMPVKEFSTDKLLEGLAINNREFIDLCILCGDYCESIK 239
QY 241 GIGGOTALKLIHQHSIESILENKDRIQIPEDWPVQEARLKEPNVT-LDIPELKWT 299
DB 240 GIGPKRALELITVTDIETILDNDSSKYTPVENWYKVALELFIPEVALADAFIDLKVV 299
QY 300 APDEGLSFLVKGNGFNEADVTKAIFKISAKNKSQGLSEFEK----PTATTSAPL 354
DB 300 EPDEGLVKFLCGDRQFNEVRNGAKLMSKQAQTVRLDSFFKTLSTPNATNA-- 357
QY 355 KRKETSDDTSKAAANKTKAGG 376
DB 358 KRK--AEAAKKSANKKAKTSG 377
RESULT 13
O9GZ01 PRELIMINARY; PRT; 650 AA.
AC O9GZ01;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Flap endonuclease-1.
GN FEN1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Casta L.J. III, Schmutte C., Taraschi T.F.;
RT "Flap Endonuclease-1 (Plasmodium falciparum).";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF278764; AAG01445.1; -
DR HSSP; Q58839; 1A76.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR000513; Exo_N_I.
DR InterPro; IPR003584; HHH_2.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF01367; 5_3_exonuclease; 1.

```

```
DR Pfam: PF00867; XPG_I; 1.
DR PRINTS: PR00853; XPGRADSUPER.
DR SMART: SM00475; 53EXOC; 1.
DR SMART: SM00279; Hh2; 1.
DR SMART: SM00484; XPGI; 1.
DR SMART: SM00485; XPGN; 1.
KW Endonuclease.
SQ SEQUENCE 650 AA; 73989 MW; 2E9464DAB7F45B1D CRC64;

Query Match 45.7%; Score 887; DB 5; Length 650;
Best Local Similarity 46.6%; Pred. NO. 1.9e-54;
Matches 184; Conservative 74; Mismatches 117; Indels 20; Gaps 5;

QY 1 MGKIGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIVVGTGM-ETLTNEAG 59
DB 1 MGKIGLTKFTIADAPNAIKIEIKIESLMGRIIADASMSLYQFIIRDSQYGNLTNEG 60

QY 60 EYVSHLQGFNRTIRLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVG 119
DB 61 EYVSHLQGFNRTIRLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVG 120

QY 120 DKDAIEKLSKRTVKVTRHNEDECKRLRLMGVPVVEAPSEAEACALCINDKVFVASE 179
DB 121 NLEETIKKSGRTVTRKONEEAKLLTLMGPIIEAPCEAESQCAFLLTKYNLAHATATE 180

QY 180 DMSLTFGAPFLRH------MDPSSKIPVMEFDVAKVLELELTMDQFIDLC 228
DB 181 DADALVFGTKLIRLNANATSNQNNKNSKRGYILTEINLEQVLKGLNLTMDQFIDFC 240

QY 229 ILGCGDYCDTGKIGGQYALKIRHSGSIESLENLKNDRYQIPEDWPYQEARLKEPN 288
DB 241 ILGCGDYCDTGKIGGQYALKIRHSGSIESLENLKNDRYQIPEDWPYQEARLKEPN 300

QY 289 VTLDIP---ELKWTAPDEGLISFLVKGDFNEDRVTKAIEIKSAKNSQGRLESFF 344
DB 301 V---LPKEDIKIDWNEPOIEELKHLKDYFNELRVNINLLKARKVTTQRRLDNEF 357

QY 345 KPTATTSAPLKRKET-SDKTSKAAANKTKAGKK 378
DB 358 TACTKTKSLIVEETKKEQTLTPARKKKRPTAGDK 392

RESULT 14
Q9U0K1 PRELIMINARY; PRT; 672 AA.
AC Q9U0K1.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Flap endonuclease 1 (Putative flap endonuclease).
GN FEN1 OR MAL4P2.21.
OS Plasmodium falciparum, and
OC Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839, 36329;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=P.falciparum; STRAIN=3D7A;
RA Li J.-L.;
RT "Characterization of the flap endonuclease 1 homolog from the human
RT malaria parasite Plasmodium falciparum."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=P.falciparum (isolate 3D7);
RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,
RA Quail M., Barrell B.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF093702; AAG37989.1;
DR EMBL: AL035475; CAB62862.1;
DR HSP: Q58839; 1A76.
DR InterPro: IPR002421; 5_3_exonuclease.
```

```
DR InterPro: IPR000513; Exo_N_I.
DR InterPro: IPR003584; HHL2.
DR Pfam: PF01367; 5_3_exonuclease; 1.
DR Pfam: PF00867; XPG_I; 1.
DR Pfam: PF00752; XPG_N; 1.
DR PRINTS: PR00853; XPGRADSUPER.
DR SMART: SM00475; 53EXOC; 1.
DR SMART: SM00279; Hh2; 1.
DR SMART: SM00484; XPGI; 1.
DR SMART: SM00485; XPGN; 1.
KW Endonuclease; Exonuclease.
SQ SEQUENCE 672 AA; 76680 MW; 6C2CDBB95F40042 CRC64;

Query Match 45.7%; Score 887; DB 5; Length 672;
Best Local Similarity 46.6%; Pred. NO. 2e-54;
Matches 184; Conservative 74; Mismatches 117; Indels 20; Gaps 5;

QY 1 MGKIGLTKLLADNAPKAMKEQKESYFGRKIAVDASMSIYQFLIVVGTGM-ETLTNEAG 59
DB 1 MGKIGLTKFTIADAPNAIKIEIKIESLMGRIIADASMSLYQFIIRDSQYGNLTNEG 60

QY 60 EYVSHLQGFNRTIRLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVG 119
DB 61 EYVSHLQGFNRTIRLEAGIKPVYVFDGKPPDMKKQELAKRYSKRDDATKDLTEAVEVG 120

QY 120 DKDAIEKLSKRTVKVTRHNEDECKRLRLMGVPVVEAPSEAEACALCINDKVFVASE 179
DB 121 NLEETIKKSGRTVTRKONEEAKLLTLMGPIIEAPCEAESQCAFLLTKYNLAHATATE 180

QY 180 DMSLTFGAPFLRH------MDPSSKIPVMEFDVAKVLELELTMDQFIDLC 228
DB 181 DADALVFGTKLIRLNANATSNQNNKNSKRGYILTEINLEQVLKGLNLTMDQFIDFC 240

QY 229 ILGCGDYCDTGKIGGQYALKIRHSGSIESLENLKNDRYQIPEDWPYQEARLKEPN 288
DB 241 ILGCGDYCDTGKIGGQYALKIRHSGSIESLENLKNDRYQIPEDWPYQEARLKEPN 300

QY 289 VTLDIP---ELKWTAPDEGLISFLVKGDFNEDRVTKAIEIKSAKNSQGRLESFF 344
DB 301 V---LPKEDIKIDWNEPOIEELKHLKDYFNELRVNINLLKARKVTTQRRLDNEF 357

QY 345 KPTATTSAPLKRKET-SDKTSKAAANKTKAGKK 378
DB 358 TACTKTKSLIVEETKKEQTLTPARKKKRPTAGDK 392

RESULT 15
Q8SS91 PRELIMINARY; PRT; 345 AA.
AC Q8SS91.
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Structure-specific endonuclease of the XPG/RAD2 family.
GN ECU03_1080.
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prensier G., Barbe V., Peyretillade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
```

RT Encephalitozoon cuniculi";  
RL Nature 414:450-453(2001).  
DR EMBL: AL590443; CAD26252.1; -  
SQ SEQUENCE 345 AA; 38943 MW; B8C57D5F4B3E4F7C CRC64;

Query Match 43.48; Score 841.5; DB 5; Length 345;  
Best Local Similarity 48.98; Pred. No. 1.4e-51;  
Matches 170; Conservative 70; Mismatches 95; Indels 13; Gaps 5;

QY 1 MGKIGLTKLLADNAPKAMKEQKFESYFGRKIADVADSMISYQFLIVVGRGTMETLTNEAGE 60  
Db 1 MGKIQLSKLLRENSKRGIRERPLVYSSKKVAIDASMSYQFLIAV-RSGGATLGN-DS 58  
QY 61 VTSHLCMFNRTIRLEAGIKPVYVFDGKPPDMKQELAKRYSKRDDATKDLTEAVEVGD 120  
Db 59 PTHLVGFFYRTIRMVDELGITPVYVFDGVPPEIKMKELEKERRAAADREVREASEVGD 118  
QY 121 KDAIEKLSKRTVKVTRHNEPCKRLLRMGVVPVVEAPSEAECAALCINDKVFVASED 180  
Db 119 KELMEYDKRTKVTGVHVECKRLLGLMGIPETAPSEAEAYCALLCKKXVYGVATED 178  
QY 181 MDSITGAPFLEHLMDFSSKIPVMEFDYAKVLELELTMDQFIDLCILCGCDYCDISIK 240  
Db 179 MDALTEGSPVLRNFNGTQSRRLPVMENLPQILEDLSDHSEFIDLCILCGCDYCSYTLK 238  
QY 241 GIGGQALKLIRQHSIESILENLKNDRYQIPEDWYQFARRLFKEPNVTLDPPELK--- 297  
Db 239 GIGPKKALGLIKKHSIGNILKN---EDLEVPGDWRYSDAOKIF---GSLAEIGEIRDFN 292  
QY 298 --WTAPDEEGLISFLVKDNGFNEDRVTKAEIKKSAKNKSSQGRLESF 343  
Db 293 ISWDSIDRNGIVNLFVEEKGFDLERNVKGIDKLINSKKKGTOGRLDCE 340

Search completed: May 8, 2003, 21:42:47  
Job time : 61 secs

